

[illegible]

## RESULT 6

```

US-09-973-451-10
; Sequence 10, Application US/09973451
; Patent No. US20020132328A1
; GENERAL INFORMATION:
; APPLICANT: JACOBSON, Myron K.
; APPLICANT: JACOBSON, Elaine L.
; APPLICANT: AM, Jean-Christophe
; APPLICANT: LIN, Winston
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOSYDROLASE
; TITLE OF INVENTION: (PARG) ENZYMES,
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUN
; TITLE OF INVENTION: THEREWITH
; FILE REFERENCE: N1AD 201
; CURRENT APPLICATION NUMBER: US/09/973,451
; CURRENT FILING DATE: 2001-10-09
; PRIOR APPLICATION NUMBER: US/09/302,812
; PRIOR FILING DATE: 1999-04-30
; PRIOR APPLICATION NUMBER: 60/083,768
; PRIOR FILING DATE: 1998-05-01
; NUMBER OF SEQ ID NOS: 38
; SEQ ID NO 10
; LENGTH: 726
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
; FEATURE:
;
US-09-973-451-10

Query Match      8.9%; Score 463; DB 9; Length 726;
Best Local Similarity 23.9%; pred. No. 1.3e-27;
Matches 207; Conservative 123; Mismatches 278; Indels 258; Gaps 37

QY      145 YLNQHTAAMCKWNEGKHTEQLLESEPTVTLVPEQFSNANIDRSPONDDHSDTDTSEEN 204
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB      27 FAHQVPTMKRRLTEHGNNTTESKD-----PBEPKSRDVFVSSQSDSDSQEDSAEN 77
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      205 -----RDNQOFLT-TVKLANAKQ-----TTDEHAREAKSHQCKSKCHPGEDCASCOQ 252
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

QY 205 -----RDNQQLT-TVKLANAKQ-----TTEDEHAREAKSHQKCSKSPGEDCASCQ 252

```

Db      78  PEIAKEVSENCENLTETLKISNTIESLDNVTRESEHTLD--NH-----KSTEPMEE----- 125
QY      253  DEIDVVPKPLSDVSGEDVGTGSKNDKLIROBSCLGNSPPFKESEPESEPMVDVDSKNS 312
Db      126  ---DVNNKNI-----DVAINSEDEDELVEEN-----NKEVR-----DGEVQQL 163
QY      313  CQSEADEE---TSPGFDEQE---DGSSSTANKPSRFQARDADIEF-----RKR 356
Db      164  SQDLFADDOELIPIPGIMKDTTTLQDITDSEVETAQKMEMIETADSTFVGSDSKATKT 223
QY      357  YSPKGG-----GEVRLH-----FOFEGESRTGMNDLNAKLPG----- 388
Db      224  VRTSSSFLSTVGTCTAPAKGRARMTYQKELEKVIATFTEG--NUTLPDLNKNVDPRNRY 282
QY      389  ---NISSLNVECRNKHQKQKODKITDHLMLRPLKAEADRRKEQWETKHQRT-----RX 438
Db      283  CTIPNFPASQGLREDNRYPGK-----IVLP-----QWREFDSKRRRDSYFYFKRK 330
QY      439  IPKVPPHLSGDKKLGTPTEEMRM--PRCGIRLPLRPSANHTVIRVDLLRAGEVPEK 497
Db      331  LDGKLCYKTTGYFMFVGLHNNWEPDPDITYKLPALE----- 368
QY      498  FPHYKDLMDNKHVKWPCSEONLYPVEDENGERTAGSRWELI--OTALLNKFTREPONLKDA 556
Db      369  ---MYKEM-----SELVGRREVELEKFAVARIA-- 394
QY      557  ILKYNVAYSKKDFTALIDFWDKVLBEAAQHLYQSILPDMVKIALCLPNICTPQIPLLK 616
Db      395  -----KTAEDILPERIYR--LVGDV----- 412
QY      617  QKMHISITMQEQIASILANAFCTTPRRNAKMKSEYSSYPDINFNRLPGRSSRKPEKL 676
Db      413  ---ESATLSHKQCAALVARMFF-----ARPDSPFS-----FCRLSSDKSICVEKL 455
QY      677  KTLFCYFRRVTEKKPTGLVTF--TQSLDF--PEWERCERKPLRLHVT--YEGTIEENGQ 731
Db      456  KFLFTFYDKMSMDPPDGAVSFRUTKMDKOTFNBEWK--DKKUSLPEVEFDEMLIEDTA 513
QY      732  GMLQVDFANRFVGGVTSAGLAVQEIIRFLINPELLIISRLFTVLDHNECHLIITGBOYSE 791
Db      514  LCTQVDFANEHLGGVLNHSQVEEIRFLMCPENMGMLICERKMKQLEAISVGVYFESS 573
QY      792  YTCYAEITYR-----SRSHEDGE--RDDCERRCTEIVADALHRR-----YLDQFVPEK 840
Db      574  YTCYGHTLKWAELQPNHRSRQNTNEFRDFGRLAVETIADAILFKGSKLDCQTEQLNKAN 633
QY      841  MRRELKAYCGFLRPGVSSNLSAVATGNWCGGAFGADARLKALIQILAAAAAERDVVYF 900
Db      634  IIREMKKASIGFWSQGPFTNI--PIVTGWWGCGAFNGDKPLKFLIIQVIAAGVADRPLHFC 692
QY      901  TFGDSELMRDYISMHIPLTERKLTUVG 926
Db      693  SFGEPELAACKKILIERMKQDVTLG 718

```

```

RESULT 7
US-10-424-599-184988
; Sequence 184988, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated with
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 184988
; LENGTH: 200
; TYPE: PRT

```

```

; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(200)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_138059C.1.pap
US-10-424-599-184988

Query Match      4.2%; Score 217.5; DB 12; Length 200;
Best Local Similarity 29.9%; Pred. No. 6.2e-09;
Matches 53; Conservative 27; Mismatches 68; Indels 29; Gaps 4;

QY      625  MQEQIATILANAFCTTPRRNAKMKSEYSSYP-----DINFNRLPGRSSRKPEK 675
Db      33  LTQETSAPPSPSRTLFCFLP-----VSDRPGIHLPMINSDVSSGLYEDYSQPEST 83
QY      676  LKTLFCYFRRVTEKKPTGLVTFTRQSLE-----DFPE--WERCERKPLRLHVTYEG 724
Db      84  TGRIAHYGQRISEMPEKGVISPERKVLFPKNDISIHSYDANFWSSTAIPLCRFEVHSSG 143
QY      725  TIEENGQMLQVDFANRFVGGVTSAGLAVQEIIRFLINPELLIISRLFTVLDHNECL 781
Db      144  LIEDQSSGAAEVDLANKYLGGALGRGCVQEIIRFVWSPLEAGMLFLPAMADNEAI 200

RESULT 8
US-10-424-599-156445
; Sequence 156445, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated with
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 156445
; LENGTH: 180
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_112291C.1.pap
US-10-424-599-156445

Query Match      3.6%; Score 188; DB 12; Length 180;
Best Local Similarity 30.1%; Pred. No. 1.2e-06;
Matches 49; Conservative 23; Mismatches 41; Indels 50; Gaps 4;

QY      844  ELNKAYCGFL-----RP----- 855
Db      7  EINKAFGFLYQCKYQYQKILQENGCTSAFYAATSTSMETDEGEISNHNKITSQNDYH 66
QY      856  GVSSENLAVATGNWCGGAFGADARLKALIQILAAAAAERD-VVYTFGDSSELMRDYISM 914
Db      67  GMDQGNIGVATGNWCGGAFGADARLKALIQILAAAAAERD-VVYTFGDSSELMRDYISM 914
QY      915  HIFLTERKLTUVGVYKLLRLRYNEECRNCSTPGPDIKLYPFYI 957
Db      126  AHWILSQRTVGDMLMNNLIEYSINRSKGETNVTGFLQWLPSIY 168

RESULT 9
US-10-424-599-233915
; Sequence 233915, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K

```



```
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 233915
; LENGTH: 300
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_5324C.1.pep
US-10-424-599-233915

Query Match          3.5%; Score 179.5; DB 12; Length 300;
Best Local Similarity 19.0%; Pred. No. 1.2e-05;
Matches 56; Conservative 45; Mismatches 65; Indels 129; Gaps 5;

QY 576 FWDKVLBEAAQHLYQSILPDMVKIALCLPNT-----CTQPIPLL 615
Db 76 FFEVMSGESSKWFQEVLPALGNLLRLPSLLESHYQNTDNMAIDGEGAMLTTALRL 135
QY 616 KOXNHSHISMQEQATSLANAFCTFPRNAKMKSEYSYP-----DINFNRLFE 666
Db 136 DSQPGIIVFLTQELIAALLSCSLFLCFP-----VSDREVIHLPMINFDVLPFGSLYD 186
QY 667 GSSRKPEKLTLCFVFRVTEKKTGLVTFTRQSLDFEPWECEKPLTRLHLVYEGTI 726
Db 187 DYSQKQENKIMCIVHYFQISEMPKGA----- 215
QY 727 BENGQMLQVDFANRFVGGVTSAGLVQBEIRFLINPELIISRLFTVLDHNECLIIIGT 786
Db 216 -----YNERIEIVGV 225
QY 787 ROYSYTGVAETVRSRSHEDGSRDDCERRCTEIVADIALHFRYLDQFVPEKM 841
Db 226 ERFSGYTHASSFRFSRAKAEGRDEDPVGRKRTSDLS-----EKYFPFKM 270

RESULT 10
US-09-864-761-43244
; Sequence 43244, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Aeomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
```

```
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annonax Sequence Listing Engine vers. 1.1
; SEQ ID NO 43244
; LENGTH: 748
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC004886.1
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.1
; OTHER INFORMATION: EST_HUMAN HIT: A1636014.1, EVALUATE 3.00e-51
; OTHER INFORMATION: SWISSPROT HIT: P38254, EVALUATE 1.40e+00
US-09-864-761-43244

Query Match          3.3%; Score 173.5; DB 9; Length 748;
Best Local Similarity 20.0%; Pred. No. 0.00015;
Matches 133; Conservative 117; Mismatches 233; Indels 177; Gaps 30;

QY 52 PSSPACVPGQAGHRGSAISLVFKQKTTTSMWDT-----KGIK--TAESSLDSKENNN 103
Db 15 PKSPQGLSDTGYSSDGISSSL---GEIPSLIPTDDEKDLKGLKXDSFSQESSPSPSDDL 70
QY 104 TRIESMWSVQKDNFYQHNVEKLVNVSQSLDKSLTEKSTQVLYNQHOTAAMCKWQNEGHK 163
Db 71 AKLESTVLISLE-----AQASTLADEKS--EKTKQ---PHEVSP--EQFQDQEK 112
QY 164 TEOLLESEPTVTLVPEQFSNANIDR-----SPQNDHSDTDSSENRDNQOFLTIVK-- 215
Db 113 TQSL--SETLETISEEIKESQERKDTFKKDSQDIPSSKHKEKSEFVDDITRRREP 170
QY 216 LANAKOTTDEHA-----REAKSHQKCSKCHPGECDASCQDQDEIDVVPKSPLSVDGSEDV 271
Db 171 YDSVEESSESNSPVPQKRRTSVGSSSSDEYKQSDSQSGSEEDFIRKQ--IENSADSD 229
QY 272 GTGSKNDNKLIRQESCLGNSPPFEKESRPE-----SPMDVDNSKNSC 313
Db 230 ASGSEDDDFIRNQLKIEISSSTESQKRETKGKGTAGKRRLTRKSTSIDEDAGRHS 289
QY 314 QDSEADETSPCFDEQEDGSSSQTANKPS-----RFQARDADIEFRKRYSTKG 361
Db 290 WHDEDE-----AFDSSPELVKRETKSQSEELVVTGGGLRRFKTIELNSTIADKYSAS 345
QY 362 GEVRLHFQF-----EGGSRT---GMNDLNK-----LPGNISLNVCEKRS 400
Db 346 SQKTSVLVFDPEPELEMESLTDSPEDSRGEGSSLSHASSFTFGTSPTSVSLSDESDSS 405
QY 401 KQHGKDKSK-----ITDHLMLPKAEDR-----RKEQWETKHQTERKI----- 439
Db 406 PSKKGESQQRKARHPHGLPLPTIEDSSSEEEELKEEELKEQKREIEQOQKSSS 465
QY 440 -----PKYVPPHLSPPDKKVLGTPPIEMRRNPRCGIRLPLLRPS-ANH 480
Db 466 KSKSKDKDELRAQRREERFKTTPSNLSPLED---ASPTTELQAAEME---ELHRSKCSY 520
QY 481 TWTIRVDLLRAGEVPKPFPTHYKDWLW-NKHKVMPCEQONLYPVEDENGERTAGSRWELI 539
```

```

Db 521 SPISIED-----PRGFISPEKIIIEVKVYKLP-TAVSLYSPTDEQ-----560
QY 540 QVALNKFTRPQNLKDAILKYNVAYSKKWDFTAL--IDFWDKVLAEAE-----AQHLY 590
Db 561 --SIMKEGSKALKSAEEMHMKTHYKAFPAANERDEVFEKEPLYGGMLIEDIY 618
QY 591 QSLIPD 596
Db 619 ESLVED 624

RESULT 11
US-09-973-451-11
; Sequence 11, Application US/09973451
; Patent No. US20020132328A1
; GENERAL INFORMATION:
; APPLICANT: JACOBSON, MYRON K.
; APPLICANT: JACOBSON, Elaine L.
; APPLICANT: AM, Jean-Christophe
; APPLICANT: LIN, Winston
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE
; TITLE OF INVENTION: (PARG) ENZYMES,
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV
; TITLE OF INVENTION: THEREWITH
; FILE REFERENCE: NIAD 201
; CURRENT APPLICATION NUMBER: US/09/973,451
; CURRENT FILING DATE: 2001-10-09
; PRIOR APPLICATION NUMBER: US/09/302,812
; PRIOR FILING DATE: 1999-04-30
; PRIOR APPLICATION NUMBER: 60/083,768
; PRIOR FILING DATE: 1998-05-01
; NUMBER OF SEQ ID NOS: 38
; SEQ ID NO 11
; LENGTH: 31
; TYPE: PRT
; ORGANISM: Bos taurus
; FEATURE:
US-09-973-451-11

Query Match 3.2%; Score 167; DB 9; Length 31;
Best Local Similarity 100.0%; Pred. No. 3.6e-06;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 770 LFTVELDHNECLIITGTEQYSEYTGIAETVR 800
Db 1 LFTVELDHNECLIITGTEQYSEYTGIAETVR 31

RESULT 12
US-10-221-278-197
; Sequence 197, Application US/10221278
; Publication No. US20040034208A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc
; TITLE OF INVENTION: No. US20040034208A1el Nucleic Acids and Polypeptides
; FILE REFERENCE: 21272-045
; CURRENT APPLICATION NUMBER: US/10/221,278
; CURRENT FILING DATE: 2002-09-06
; PRIOR APPLICATION NUMBER: 09/693,267
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 09/665,363
; PRIOR FILING DATE: 2000-09-19
; PRIOR APPLICATION NUMBER: 09/616,847
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: 09/596,193
; PRIOR FILING DATE: 2000-06-17
; PRIOR APPLICATION NUMBER: 09/574,454
; PRIOR FILING DATE: 2000-05-19
; PRIOR APPLICATION NUMBER: 09/519,705
; PRIOR FILING DATE: 2000-03-07
; NUMBER OF SEQ ID NOS: 752
; SEQ ID NO 197
; LENGTH: 1163

```

```

; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-221-278-197

Query Match 3.2%; Score 166.5; DB 12; Length 1163;
Best Local Similarity 20.0%; Pred. No. 0.0011;
Matches 161; Conservative 109; Mismatches 299; Indels 235; Gaps 38;

QY 19 ATTSPAASDARS-----PESRQRVLDPKDAHVQFRVPPSSPACVPGQAGQH-----65
Db 186 AVSSLNSHRSRSHGNDHHSKHQSRKPRDPDANW-----DSPSRVPFSSQGHSTQSPPPS 241
QY 66 -RGSATSLVFKQKTIITSMWTGKGTAE-----SESILDSKENNNNTRIESMMSSVQKDNFYQ 120
Db 242 LMSKNSMLQKPTAYVRPMD--GQESMEPKLSSEHYSSQSHGNSMTELKPS-----X 292
QY 121 HNVEKLNVVQSILDKS-----LTEKSTQYLNQHQTA--AMCKWQNE-----GKH 163
Db 293 AHLTKLKIPSQ-PLDASASGVDCVDEILKEMTHSWPPPLTAHTFCKTEPSEKFPPTKE 351
QY 164 TEQLLESEPTVTLVPEQFSNANIDRSPQND-----HSPDTSEENRDNQOFLTIVKLAN 218
Db 352 SQGSNFTGEQKRYNPKTSNGHQSKMLKDDLLKSSSDSDGEQCDK-----TMPSTP 407
QY 219 AKQTTDEHAREAKSHQKCSKCHPGEDCASCQODEIDVVVFKPSLSDVGSDEVTGSGKND 278
Db 408 GSNSEPSHNSGADNSRDDSSSHSGSESS-----GSDSESESSSD 450
QY 279 NKLIROESCLGNSPPFEKESEPEP-----MDVDN-----308
Db 451 SE-----ANEPQSASPEPEPTTNKWLQNLNKNVPHKVPASSVDSNIPSQGY 502
QY 309 -----SKNSQDSEADBEYS---PGHEQ--EDGSSSQATANKPSRFQADADIEFRK 355
Db 503 KKEGREQGTGNSYDTSGPKTSSATPGRDSKTIQKSGESGRGRQKSPQS--DSTQRT 561
QY 356 --RYSTKGGEVRLHFQFEGG---ESRTGMNDLNAKLPGNISSLNVECRNKGQHGKDKSI 410
Db 562 VGKKQPKKAEKAAAEPRGGLKTESETPV-DLASSMPSSRRHKAATK-GSRKPNIKKESKS 619
QY 411 TDHMLPLKPADRRKQWETKHORTERKI-----PKYVPPHLS 448
Db 620 SPR-----PTAE-KKTKYKTSKSKQKREIITETDTSSSDSESLPPSSQTPKIPESNRT 674
QY 449 PDK-----KWLGTPIEEMRMPCGIRLPLLRPSANHTVTRVDDLEAGVVP- 495
Db 675 PVRPSSVEEDSFPRQMFSPMEKEKLS-----PLSEPDYPLVLIKIDLLNLTPIG 728
QY 496 KPFPFTHYKOLMNDKHKVMPCEQNLYPVEDENGERTAGSRWELIQTALLNKFTRPQNLKD 555
Db 729 KP-----YKETEPPKGEKNVPEKHTREAAQKASEKVS-----NKGKREKHNED 772
QY 556 AILKYNVAYSKKWDFTALIDFWKVLAEAEQHL-----YQSLIPDMVKIALCLPNICTQP 611
Db 773 ---DNRASEKK-----PKTEDKNSAGHKFSSNRESSQSAKAKEKDLLPS-PAGP 818
QY 612 IPLLKQNMHSI---TMSQEQIASILLANAFCTFPRRNAMKSEYSSYPDINFNRLPEGR 668
Db 819 VPKDPKTEHSGSKRTISQSSSLKSSNS-----NKTSGSSKNSSSTSKQKTEGK 870
QY 669 SSRKPEKIKTLFCYFRVRVTEKKPT 692
Db 871 TSSSSSEVK-----EKAPS 884

RESULT 13
US-10-291-172-197
; Sequence 197, Application US/10291172
; Publication No. US20030228584A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc
; TITLE OF INVENTION: No. US20030228584A1el Nucleic Acids and Polypeptides
; FILE REFERENCE: 21272-045

```

```

; CURRENT APPLICATION NUMBER: US/10/291,172
; CURRENT FILING DATE: 2000-11-08
; PRIOR APPLICATION NUMBER: 09/693,267
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 09/665,363
; PRIOR FILING DATE: 2000-09-19
; PRIOR APPLICATION NUMBER: 09/616,847
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: 09/596,193
; PRIOR FILING DATE: 2000-06-17
; PRIOR APPLICATION NUMBER: 09/574,454
; PRIOR FILING DATE: 2000-05-19
; PRIOR APPLICATION NUMBER: 09/519,705
; PRIOR FILING DATE: 2000-03-07
; NUMBER OF SEQ ID NOS: 752
; SEQ ID NO 197
; LENGTH: 1163
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-291-172-197

Query Match      3.2%; Score 166.5; DB 15; Length 1163;
Best Local Similarity 20.0%; Pred. No. 0.0011;
Matches 161; Conservative 109; Mismatches 299; Indels 235; Gaps 38;

QY 19 ATTSPAADARS-----PFSRQRVLDPKDAHQVFPVPPSPACVPGQAGQH-----65
Db 186 AVSLSNSHSHRGNDHSHKQSKSPDPDANW-----DSPRVFPSSGQHSQTSPFPS 241
QY 66 RGSATSLVPKOKTITSMWDTGKIATB-----SESLSKENNNNTLESMMSSVKQDNFYQ 120
Db 242 LMSKSNMLQKPTAYVRPMD--GOESMEPKLSSEHYSSQSHGNSMTLKPSS-----K 292
QY 121 HNVEKLVNVSQSLDKS-----LPEKSTQVYNQHOTR--AMCKWONE-----GKH 163
Db 293 AHLTKLKPQSQ--PLDASGDNVSCVDEILKEMTHSWPPPLTAIHFPCKTEPSKFPPTKE 351
QY 164 TEQLLESEPQTVTLVPEQFSNANIDRSPQND-----HSDTSENRDNOQFLTTVKLAN 218
Db 352 SQSNFGTGEQKRYNFKTSNGHQSMLKDLKLSSSDSDGQDCDK-----TPRSTP 407
QY 219 AKQTTDEHARAKSHQKSKCHPQEDCASQOQDEIDVVPKSPISDVGSEDVGTGSKND 278
Db 408 GSNSEPPSHNSGADNSRDDSSSHSGSESS-----GSDSESSSSSD 450
QY 279 NKLIROESCLGNSPPPEKSEPEP-----MDVDN-----308
Db 451 SE-----ANEPSOSASPEPEPPTNKWLDNWLKNVPHKVPSPASSVDNIPSSQGY 502
QY 309 -----SKNSQDSEADEETS---PGFDEQ--EDGSSSQTKANKSPRFOARDADIEFRK 355
Db 503 KKEGRQGTGNSYTDTSQPKETSSATPGRDSKTIQKSESGRGRQKSPAQSDSTTQRET 561
QY 356 ---RYSTKGGEVRLHFQEGG---ESTGMNDLNAKLPGNISLNVCRNSKQHGKDSKI 410
Db 562 VGKQPKQAKAAAEPRGLKTIKTESETPV--DLASSMPPSRHKKAATK--GSRKNIKESKS 619
QY 411 TDHLMRLPKAEDRRKQWETKHQTERKI-----PRYVPPHLS 448
Db 620 SPR-----PTAE--KCKYKTSKSKQKREIETDTSDDSDSESLSPPSQTPKYPESNRT 674
QY 449 PDK-----KWLGTPIEMRRMPCGIRLPLLRPSANHTVIRVLDLRAGEVP--495
Db 675 FVKPSSVEEDSFPRQMPFMEKEKLS-----PLSEPDPRYPLIVKIDLNLTRIP 728
QY 496 KPPPTHYKDLWNKHVQMPSEQNLYPVEDENGERTAGSRWELIQTALLNKETRPQNLKD 555
Db 729 KP-----YKTEPPKGEKKNVPEKHTRAQKQAEKVS-----NKGKRRKHND 772
QY 556 AILKYNVAYSKWDFDTALIDFWDKVLEAEQAHL-----YQSILPDMVKIALCLPNTCTOP 611
Db 773 ---DNRASEKK-----PKTEDNSAGHKPSSNRRESSKQSAKEKDLPS--PAGE 818

```

```

QY 612 IPLLKQKWHSI---TMSQEQIASLLANAFCTFPRRNAMKMSYSSYPDINFNRLPEGR 668
Db 819 VPSKOPKTEGSRKRTITSSQSSLSKSSNS-----NKETSGSSKNSSSTSKQKTEGK 870
QY 669 SSRKPEKLKTLFCYFRFVTEKKPT 692
Db 871 TSSSSKEVK-----EKAPS 884

RESULT 14
US-10-282-122A-43837
; Sequence 43837, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282.122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 43837
; LENGTH: 1274
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-10-282-122A-43837

Query Match      3.2%; Score 165.5; DB 12; Length 1274;
Best Local Similarity 20.3%; Pred. No. 0.0015;
Matches 104; Conservative 75; Mismatches 173; Indels 161; Gaps 20;

QY 49 RYPPSPSA-----CVPGQAGQHR-----GSATSLVPKOKTITSMWDTGKIATA 91
Db 203 RIVPSKPSKVESDKQYDYVAKTQTSQNKOLEQKQNDSDVVKQGTASKSDENVSSIT 262
QY 92 ESESLDSKENNNTRIESMMSS-----VQK-----115
Db 263 KSMFNYSKVDNITKNIENIVASQIVEIRERERKVLQKRRFKKALQKREHKNEBQDAI 322
QY 116 ----DNFYQHNVEKLVNVSQSLDKSLTEKSTQYLNQHOTAAAMCKWQNEGKHTQLESE 171

```



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 26, 2004, 18:23:35 ; Search time 26.8489 Seconds  
(without alignments)  
3496.721 Million cell updates/sec

Title: ' US-09-302-812-4  
Perfect score: 5190  
Sequence: 1 MNAGGCEPCTKATRWGAAT.....YHAVSCAETADHSQRTGT 976

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR 78: \*  
1: pir1: \*  
2: pir2: \*  
3: pir3: \*  
4: pir4: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	511	9.8	997	B84726	probable poly(ADP-ribose)
2	463	8.9	726	T21138	hypothetical prote
3	340	6.6	364	A84726	probable poly(ADP-ribose)
4	180	3.5	1298	I54367	X-linked nuclear p
5	178	3.4	1359	T34036	hypothetical prote
6	168	3.2	3147	T18674	hypothetical prote
7	166.5	3.2	1641	I38614	helicase II - huma
8	165.5	3.2	1274	A89959	hypothetical prote
9	165	3.2	884	D96730	unknown protein F5
10	164	3.2	4717	T41581	hypothetical coile
11	161	3.1	1788	T29043	hypothetical prote
12	159.5	3.1	1183	S65236	probable membrane
13	159	3.1	406	S38170	SRP40 protein - ye
14	159	3.1	1403	A47328	natural killer cel
15	159	3.1	2429	SJH28	spectrin alpha cha
16	156.5	3.0	1538	T29095	cardiac muscle fac
17	155.5	3.0	1974	T30010	hypothetical prote
18	154.5	3.0	646	S15901	chromogranin B pre
19	154.5	3.0	3225	I52300	giantin - human
20	154.5	3.0	3259	A56539	giantin - human
21	154	3.0	1938	JC5421	smooth muscle myos
22	154	3.0	1972	JC5420	smooth muscle myos
23	153	2.9	1031	T38411	probable GTPase ac
24	153	2.9	1071	T85343	hypothetical prote
25	153	2.9	6642	T29757	protein UNC-89 - C
26	152	2.9	1320	S57113	BUD4 protein - yea
27	151	2.9	1272	T30593	hypothetical prote
28	150.5	2.9	1312	1	RAD50 protein - ye
29	150.5	2.9	1957	2	hypothetical colle

30 150.5 2.9 1972 1 A41604 myosin heavy chain  
31 150 2.9 1206 2 T34021 protein kinase SK2  
32 150 2.9 2137 1 SJHUB spectrin beta  
33 149.5 2.9 1210 2 I39410 AF-4 protein, spli  
34 148.5 2.9 472 2 T27903 hypothetical prote  
35 148.5 2.9 665 2 B71609 hypothetical prote  
36 148.5 2.9 677 1 CNHUB chromogranin B pre  
37 148.5 2.9 1928 2 S46773 myosin heavy chain  
38 148 2.9 644 2 S55395 neurofilament prot  
39 148 2.9 1271 2 A45555 glutamate rich pro  
40 148 2.9 1780 2 T17272 hypothetical prote  
41 147.5 2.8 2218 2 B84683 hypothetical prote  
42 146.5 2.8 667 2 A40713 cylindric I - bovine  
43 146.5 2.8 867 2 T21311 hypothetical prote  
44 146.5 2.8 1463 2 T30290 AAS surface protei  
45 146.5 2.8 2649 2 A40937 bullous pemphigoid

ALIGNMENTS

RESULT 1

B84726  
probable poly(ADP-ribose) glycohydrolase [imported] - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 02-Feb-2001  
C:Accession: B84726  
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;  
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.  
eues, D.; Niernman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.  
Nature 402, 761-768, 1999  
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.  
A:Reference number: A84420; MUID:20083487; PMID:10617197  
A:Accession: B84726  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-997 <STO>  
A:Cross-references: GB:AE002093; NID:g4887749; PIDN:AAD32285.1; GSPDB:GN00139  
C:Genetics:  
A:Gene: At2g31870  
A:Map position: 2

Query Match 9.8%; Score 511; DB 2; Length 997;  
Best Local Similarity 31.8%; Pred. No. 2.4e-23;  
Matches 135; Conservative 53; Mismatches 118; Indels 114; Gaps 12;  
QY 577 WDKVLEEAQHLVQSILPDMVKIALCLPNI-----CTQPIPLKQKXNHS 622  
DB 79 FELIDEKESKWFDEIIIPALASLLLOFFSLLEVHFQADNIVSGIKTGLRLNSQQAGI 138  
QY 623 ITMSQEQIASLLANAFCTFPRNNAKMKSEYSPDINFNR--LFGRSRKPEKLTFLF 680  
DB 139 VFLSQELIGALLACSPFCLFPDNRGAK---HLPVINFHLSLYSYSQSQSKIRCI 194  
QY 681 CYFRRVTEKPTGLVTFTRQ--SLEDFFPWERCEKELTRLVHTVEGTIEHQGMQVDF 738  
DB 195 HYFERFCSCVPIGVSFERKITAAPADFWSKSDVSL-----YQ-----PDALEVD 242  
QY 739 ANRFVGGVTSAGLVOEIRFLINPELIIISRLFTFVLDHNECLIIITGQSYTYGVAET 798  
DB 243 ANKYLGGSLRGCVQVEERFMINPELIAGMLFLPRMDDNEAIEIVGAERFSCTGYASS 302  
QY 799 YRWSRSHEDGSRDDCERCCTEIVADAL-----HFRYLDQFVPEKRRRLNKAYCGF 852  
DB 303 FRPAGEYIDKAMPKRRTRIVAIDALCTPQRHFQDIC-----LLREINKALCGF 355  
QY 853 LR-----854  
DB 356 LNCCKAWEHQNIIFMDGSDNEIQLVRNGRSGLLRTTETTSHTPLNDVEMNRKPNLLI 415  
QY 855 -----PQVSSENI--SAVATGNWCGAFGGDARLKALIQILAAAAARVDVV-YFTFGDSE 906  
DB 416 RDPFVGVGDNEHDDGVATGNWCGVFGGDPKATIQWLAAASQTRPFIYSYVTFG-VE 474

QY 907 LMRDI 911  
 Db 475 ALRNL 479

RESULT 2  
 T21138  
 hypothetical protein F20C5.1 - Caenorhabditis elegans  
 C:Species: Caenorhabditis elegans  
 C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 04-Mar-2000  
 C:Accession: T21138  
 R:Matthews, P.  
 submitted to the EMBL Data Library, December 1995  
 A:Reference number: Z19381  
 A:Accession: T21138  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-726 <WIL>  
 A:Cross-references: EMBL:Z68161; PIDN:CAA92299.1; GSPDB:GN00022; CESP:F20C5.1  
 A:Experimental source: clone F20C5  
 C:Genetics:  
 A:Gene: CESP:F20C5.1  
 A:Map position: 4  
 Query Match 8.9%; Score 463; DB 2; Length 726;  
 Best Local Similarity 23.9%; Pred. No. 1.3e-20;  
 Matches 207; Conservative 123; Mismatches 278; Indels 258; Gaps 37;  
 Db 145 YLNQHTAMQWQNEGHKHTTOLLESEPTVTLVPEQFSNANIDBSQNDHSDHSDSEEN 204  
 Db 27 FAHQVPTMKRRKLTGHTNTTESKED-----PEPKSRDVFVSSQSDSEQDSASN 77  
 QY 205 -----RDNQOFLT-TVKLANAKQ-----TTDEHAREAKSHOKSKCHGEDCASQ 252  
 Db 78 PEIAKEVSENLTETLIKISIESLQNVTERSEHTLD--NH-----KSTPEWE----- 125  
 QY 253 DEIDVVPKSPISDVGEVDVGTGKNDKILIRQESCLGNSPPFEKSEPESPMDVNDKNS 312  
 Db 126 ---DVNKSNI-----DVAINSEDEDELVALEN-----NKEMR-----DGEVQQL 163  
 QY 313 QDSRADEE---TSQGFDEQ-----DGSSTQANKSRQARDADIEF-----RKR 356  
 Db 164 SQDLFADDDQELIEYFGIMKDTTQDITDSEVETAQKMEMIEETEADSTFVGEDSKATKT 223  
 QY 357 YSTKG-----GBVRLH-----FQFEGGSRTGMDLNKALPG----- 388  
 Db 224 VRTSSSSFLSTVSTCEAPAKGRARMYQKELEKHVIAFTG-NLTLQPLNKVDPDNYRY 282  
 QY 389 ----NISSIANVECRNSKHQKDKSKITHMLPMLPAEDRRKQWETKHQRT-----RK 438  
 Db 283 CTIPNFPASQGLREDNRYGPK-----IVLP-----QWKEFDSRGRERDSYFYFKR 330  
 QY 439 IPKYVPPHLPDKKWLGPPIEMRRM-PRCGRLPLLRPSANHTVIRVDLLRAGEVPEK 497  
 Db 331 LDGLKCYKTKTYGFVFMVGLLHNMWEEFDDITTKLPALF----- 368  
 QY 498 PPTHYKLDNDKHKVMPQCEQNLYPVEDENGERTAGSRWEALT-QTALLNKFTRPQNLKDA 556  
 Db 369 --MYKEM-----KTAEDILPERIYR-LVGDV-----SELVGREVLEKPARVARIA- 394  
 QY 557 ILKYNVAYSKWDFALIDFDWKVLEBAQHLQSIPLDMVKIALCLPNICTQIPILK 616  
 Db 395 -----KTAEDILPERIYR-LVGDV----- 412  
 QY 617 QKMNHSITMSQEQIASILLANAFCTTFFPRNAKMSYSPDINENRLEFGSSRKPKL 676  
 Db 413 ----ESATLSHKQCAALVARMFF-----ARPDSPFS-----FCRILSDSKSICVVKL 455  
 QY 677 KTLFCYFRVRVTEKPTGLVTF--TQSLRDF--PEWERCEKPLTRLHVT--YEGTIEENQ 731

Db 456 KFLTYFDKMSMDPPDGAUSFRLTKMDKOTNEWK--DKKLSRLPEVEFDEMLIEDTA 513  
 QY 732 GMLQVDFANRFVGGVTSAGLVOEBRFLINPELLISRLTEVLDHNECLLITCTEQYSE 791  
 Db 514 LCTQVDFANHLGGVNLHSGVQEBEIRFLMCPENVMGMLCEKMKQLEAISIVGAYVSS 573  
 QY 792 YTGVAETYRW-----SRSHEDGSE-RDDCERRCTEIVADALHFR-----YLDQFVPEK 840  
 Db 574 YTGCHTLKWAELQPNHSRQNTNEFRDFGLRVEITAIIDALLFKGSKLQCTEQNLKAN 633  
 QY 841 MRRELKAYCGFLRPGVSSSNLSAVATGNWGCAGFGDARLKALIQILAAAAERDVVYF 900  
 Db 634 IIREMKKASIGFMSQGPFTNI-PIVTGWWGCGAFNGDKPLKFTIIQVIAAGVADRLPHFC 692  
 QY 901 TRGDSLMRDIYSWHIFLTERKLTVG 926  
 Db 693 SPGEPELAACKKIIERMKQKQVTLG 718

RESULT 3  
 A84726  
 probable poly (ADP-ribose) glycohydrolase [imported] - Arabidopsis thaliana  
 C:Species: Arabidopsis thaliana (mouse-ear cress)  
 C:Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 02-Feb-2001  
 C:Accession: A84726  
 R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;  
 M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Unayam, L.; Tallon, L.  
 euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J  
 Nature 402, 761-768, 1999  
 A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.  
 A:Reference number: A84420; MUID:20083487; PMID:10617197  
 A:Accession: A84726  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-364 <STO>  
 A:Cross-references: GB:AE002093; NID:9487750; PIDN:AAD32286.1; GSPDB:GN00139  
 C:Genetics:  
 A:Gene: At2g31860  
 A:Map position: 2

Query Match 6.6%; Score 340; DB 2; Length 364;  
 Best Local Similarity 27.0%; Pred. No. 1.8e-13;  
 Matches 104; Conservative 58; Mismatches 119; Indels 104; Gaps 11;  
 QY 581 LEEAEQHLQSIPLDMVKIALCLPNICTQ-----PIPLKQKMNHSITMS 626  
 Db 1 MSKESSRWFEFLPAMACLLLRPFLSLESHYNSMDLNGTKGLRVLPNKAGIVPLS 60  
 QY 627 QEQIASILANAFCTTFFPRNAKMSYSSY-PDINFNRLF-----EGRSSRKPCKLTKLF 680  
 Db 61 QELIGALLSCSFFCLFP-----VDRGSHLPIINFDKLFGSLINTGRNEHQENKIKII 115  
 QY 681 CYFRVTEKPTGLVTTTQ--SLE-----DFPEWERCEKPLTRLHVTYEGTIEENQ 732  
 Db 116 HYFORLSSSIIPGVSFPERKILSLEQSDSTLDEFGWKGSTVNLCPVETVTSGLLEDQSV 175  
 QY 733 MLOYDFANRFVGGVTSAGLVOEBRFLINPELLISRLTEVLDHNECLITGTQYSEY 792  
 Db 176 ALEVDFANLGGALRGCVQEBIRFMINPELIVGMLFLPTMEVTEAIEVVGARFSLY 235  
 QY 793 TGYAETVRSRSHEDGSDRDCERRC-TEIVADALHFRYLDQFVPEKMRRELKAYCG 851  
 Db 236 TG-----CFRKAETRIVAIDALR----- 253  
 QY 852 FLRGVSSSNLSAVATGNWGCAGFGDARLKALIQILAAAAERDVVYF-----FGD 904  
 Db 254 --HFGVS-----QYKLESLVLSLSSSGRPIRLYMGVSLSLOGIGD 292  
 QY 905 SELMRDIYSWHIFLTERKLTVG 929  
 Db 293 VLMVEILLSSLLFFNGLRFRHSNLY 317



Db 401 RLEKKQKFNQVILEGEDITLMTGTSTQSKLVSV---LDPD-----SSTVDESKP- 452  
Qy 467 CGIRLPILRPSANHTVTIRVDLLRAGEVPKPPFTYKDLWDNKHVMKPCSEQNLYPVEDE 526  
Db 453 -----VEVHSLVR---ILKPHQAH-----GIQFMYDCAPE 480  
Qy 527 NGER--TAGSRWELIQTALLNKTRPQNLDAILKYNVAYSKWDFTALLDFWDKVLREA 584  
Db 481 SLDRDLDEGS-----GGILAHCMGLGKT---LQVITFLHTVL--- 514  
Qy 585 EAQHLVQSLPDMVKIALCLP-NICTQPIPLLKQNMHSITMSQEQIASLANAFCTFP 643  
Db 515 -----MHKIGEKCKRVLVVVKVNIWF-----KBFQKWLVDNDEELDTDVN----- 559  
Qy 644 RRNAKMKSESYSPDINFNR--LPEGRSSRKPEKLTFLCYFRRT-----EKQPTGLV 695  
Db 560 -----ELDSYKTIEDRRALKAWHSSKTPSVMIIIGYDLFRILTVEDDPKCKKPKNRN 611  
Qy 696 TFRQSLDF-----PEWERC-----KPLFLHVTYGTTEENG 730  
Db 612 RLLEKAKEDFRKYLQNFQPMVWVDEAHKLNDDSAKSCMWKILTKRRLCTGTPLQNN 671  
Qy 731 -----CGML--QVDFANRFVGGVTSAGLVQ-----EIRFL 760  
Db 672 LMEYHVMVNFVKPLGLGKTGFANRFV--NLIINRGTKDASPLEVSPM 717  
  
RESULT 6  
T18674  
hypothetical protein T04F3.1 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C>Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 29-Oct-1999  
C:Accession: T18674; T24464  
R:White, S.  
submitted to the EMBL Data Library, June 1996  
A:Reference number: Z19004  
A:Accession: T18674  
A>Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-3147 <WIL>  
A:Cross-references: EMBL:Z74026; PIDN:CAA98419.1; GSPDB:GNO0023; CESP:T04F3.1  
A:Experimental source: clone B0240  
R:Kershaw, J.  
submitted to the EMBL Data Library, May 1996  
A:Reference number: Z19894  
A:Accession: T24464  
A>Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-3147 <W12>  
A:Cross-references: EMBL:Z72513; PIDN:CAA96672.1; GSPDB:GNO0023; CESP:T04F3.1  
A:Experimental source: clone T04F3  
C:Genetics:  
A:Map position: 5  
A:Introns: 338/3; 417/3; 1957/2; 2358/3; 2400/1; 2529/2; 2585/1; 2700/3; 2744/1; 2877/3;  
  
Query Match 3.2%; Score 168; DB 2; Length 3147;  
Best Local Similarity 19.4%; Pred. No. 0.12;  
Matches 155; Conservative 114; Mismatches 322; Indels 208; Gaps 30;  
  
Qy 64 QHRSATSLVFK-----QKTIISMDTKGIKTASSESLDSKENNTR 105  
Db 1042 EHRTSAVIDLEKVFHIGSSKKPKNDDKIRGIAEFRTQKEAQRSTVTSQNSR 1101  
Qy 106 I--BSMSSVQKDFYQHNVEKLVNVSQSLSDKSLTEKSTOYLNQHQTAAMCKWQNEGKH 163  
Db 1102 IFESSISMDVFNLSHNSQVEITFASPSDLVLTSTTFHNVIE---KIDDVTK 1157  
Qy 164 TEQLESEPTVTLVPQFSNANIDRSPQNDHSDTSEE---NRDNOQFLLTVKL----- 216  
Db 1158 TDSNVVEEKEQVRLRIDFKPTFEOQLQKEFELTKKEEYSVMKNRNTSAVSDLDKVF 1217  
Qy 217 -ANAKQTT-----EDEHAREAKSHQK--CSKSHFGE---DCASCQQDEI- 255

Db 1218 DQSKETTVTSNETDEKIKGIAEFERSKQKEVQSGVAETSHSGKHIFESHISMDDVF 1277  
Qy 256 -----DVVPKSPLSDV3SE-----DVTGSKNDKNKLIRQSCGLNSP 292  
Db 1278 NTSQKYKSDKLSPPERTVEPVSTATMNLNIIIFASGIATREENTDVLBEERIQRKVE 1337  
Qy 293 PFKESEEP-ESPMVDVNSKNSCODSEADRETSFGFDEQDG-----SSQTANKSPRFOA 346  
Db 1338 EFKTTTENLEIQEVVLTTRKEVDNSDVKEHRTSAVNIDLDVFIQRSSKHPENDEDEKI 1397  
Qy 347 RDADIEFRKRYSTKGGEVRLHFQEGESKRTGMNDLNAKLPGNISLNVCRNSKHGKK 406  
Db 1398 RRGIAEFRTQKEK--EAGRSVAVIETSQNKHIFD-----KSNISM--DEVFNESQNGQK 1448  
Qy 407 DSKITDHLMLRPLKAEARRKEQWETGQRTERKIPKVPVPHLSDPKWLGTPIEEMRRMPR 466  
Db 1449 DSSNID-----MKETDMPKEDDQRYVDVH--RDKK-----PFENGSEPEPT 1488  
Qy 467 CGIRLPILRPSANHTVTIRVDLLRAGEVPKPPFTYKDLWDNKHVMKPCSEQ----- 518  
Db 1489 FNGSKISNFKQISITINLD-----NVFTEBPKLVAEDNCEIEAEEERIRKRIKQ 1540  
Qy 519 -----NLYPVEDENG-----RTAGSRWELIQTALLNKFTRPQ--LXDA 556  
Db 1541 FERTTGEQELTKNSEPAEDTSEKKHRTAAVSIDLDKVFQGTAKKPENDFDEKIKRG 1600  
Qy 557 ILKY-----NVAYSKKWDFTALIDFWDKVLREABAOHLYQS----- 592  
Db 1601 IAEFERSKQKEVQSGVAETSHSGKHIFESHISMDDVNTSQK---YKSDKLSPTPR 1657  
Qy 593 -ILPDMVKALCLPNICTQPIPLLKQNMHSITMSQEQIASLANAFCTFPFR----- 645  
Db 1658 TVPEFVSTATMNLNIIIFASGIATREK-NTDVLBEERIQRVEE-----FKKTENLEI 1711  
Qy 646 -----NAKMKSESYSPDINFNRLEPGRSSRK-----EKLTLFCYFRRT 687  
Db 1712 QKEVLTKEBGDSDVKDKHAKASAVNIDLDVFTQSSKHPENDEDEKIRGIAEFRTK 1771  
Qy 688 EKXPT--GLVTFTRQSLD 704  
Db 1772 QKEAQRSTVETQYSSKD 1790  
  
RESULT 7  
I38614  
helicase II - human  
C:Species: Homo sapiens (man)  
C>Date: 29-May-1998 #sequence\_revision 29-May-1998 #text\_change 05-Nov-1999  
C:Accession: I38614  
R:Stayton, C.L.; Dabovic, B.; Gulisano, M.; Gez, J.; Broccoli, V.; Giovanazzi, S.; Boss  
Hum. Mol. Genet. 3, 1957-1964, 1994  
A:Title: Cloning and characterization of a new human Xq13 gene, encoding a putative heli  
A:Reference number: I38614; MUID:95179111; PMID:7874112  
A:Accession: I38614  
A>Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-1641 <RES>  
A:Cross-references: EMBL:U09820; NID:G606832; PIDN:AAC50069.1; PID:G606833  
C:Genetics:  
A:Gene: RAD54L  
  
Query Match 3.2%; Score 166.5; DB 2; Length 1641;  
Best Local Similarity 19.5%; Pred. No. 0.061;  
Matches 206; Conservative 157; Mismatches 366; Indels 325; Gaps 56;  
  
Qy 75 KQTIISMDTKGIKTASSESLDSKEN--NNTRIESMSSVQKDNFYQHNVEKLVNVSQ 132  
Db 277 ERRLSKRNTKIQSGSSSS--DAEESSENKKKQRTSSKKKAVIVKEKRNLSRTSK 335  
Qy 133 SLDKSLTEKSTOYLNQHQTAAMCKWQNEGKHTEQLLESEPTVTLVPEQ---FSNANIDR 189  
Db 336 RKQADITSSSSDIEDDDQNSI---GEGSSDEQKI--KPVTENLVLSSHTGFCQSSGDE 389



QY	190	S-----PONDHSDTDSSENRDNOQFLTVK--LANAKQTTDEHAREAKSHQKCSKS	240
Db	390	ALSKSPVTVDDDDNDPENRIAKKMLLEEIKANLSSDEDDGSSDDPEERGKRTKQNE	449
QY	241	CHPGDCASQOQDEIDVVPKSPISDVGSREUVGTSGKND---NKLIIQESCLGNPP-PEK	296
Db	450	ENPQDEAKQVNS-----ESDSSE---SKKPRYRHLRLHKLTVSDGESGEEK	497
QY	297	ESPESEPMDV---DNSKNQSDSEADDETPGDEQDEGSSSQNTANKPSRFQARDADIEF	353
Db	498	KTRPKSEKHYKGRNRKVSSEDSSEDSFQESGVSEV--SESEDEQPRTRSAKARLEE	555
QY	354	RKR-YSTKGGEVRLHFQEGG-----ESRTGMNDLNAKLPL	387
Db	556	NQRSYKQKKRRRAIKVOEDSSSENKNSSEEEKEKEEEEEEEEEDEED-DSKSP	614
QY	388	GNISSLNVCECRNSQHQK--KDSKI-TDHLWRLPKAFDRKQEWETKHQTERKIPKYVP	444
Db	615	G-----KGRKKIRKILKDDKLRTETQNALKEEERRKRAERER--EKLREVIE	663
QY	445	PHLSPDKMWLGTPI-----EWRMRPRCIGIRPLLRPSANHTVTRVLLRAGEVPKPF	498
Db	664	IEDASPTKPIITTKVLDEDEETKE-----PLVQVHRNMVWL-----XP-	703
QY	499	PTHYKD---LWQNHVKYKMPSEONLYPVEDENGERTAGSRWELIOTALINKFTRPQNLK	554
Db	704	--HQVDGVQFMWD-----CCCSV-----KTKKSPGGCIIAHCMGLCK-----TLQ	744
QY	555	DAILKYNVAYSKKWDF-TALIDFMDKVL-----EEAEOHLYOSILPDMVKIALCLPNIC	608
Db	745	VVSLHTVLLCDKLDLFTALVGLSSSILAFNMWNEFEK--WQEGLKDDKLEB-VSELA	799
QY	609	TQPIPLLKQKNHISITVSOBQIASLANAFCTPPRRNAKKSEYSSYPDINFRLFEGR	668
Db	800	TVKRP---QERSYMLQRWQEDGGVMIIG-----YEMY-----RNLAQGR	835
QY	669	SSRPEKLTILFCYFRRVTEKPTGLTYTFTQSLDEPWECEKPL-----TRLHYTYEG	724
Db	836	NV-KSRKLKEIF--NKALVDPGDPFVVCDEGHILKN--EASAVSKAMNSTRSRRRIILT	890
QY	725	T-----TEENGQGMLOVDFAFRFV-----GGGVTS-----	749
Db	891	TPLONNLIEVHCWVFNKENLLGSIK-EFRNRFINPIONQCADSTWVYRVWKKRAHIL	949
QY	750	---AGLVQEE---IRFL-----INPELII SRLFTVLDHNECLIIITGTOYSE	791
Db	950	YEMLAGCVQRKYDTALTFLPKPKHEYVLAVMTSIOCKLYQYILDH-----LTGVGNNS	1004
QY	792	---YTGAYATYW-----SRSHEDGERDDCERCETIIVADIAHF	829
Db	1005	GGRGKAGAKLQDFQMLSRINWTHPCWLQLOYSIKENGYDEDSMD---EFTASDSE	1060
QY	830	RRYL--DOFVPEKMRRELNKAYCGFLRPGVSSENLSAVTCN-----WGCAPFGD	878
Db	1061	SMLSLSSDDYTKKKKKKKGK-----KSSSSGSGSDNDVEIKVWNSRSGGG	1108
QY	879	ARLKALIQILAAAAAERDVVYTFPGDSELMRDIIYSMHIFLTERKLTVGDVYKLLRYNE	938
Db	1109	E-----GNVDETCNNFVSVKLEESKAT-----	1131
QY	939	ECRNCSTPGPDIKLYPFIYHAVESCAETADHSQQ	972
Db	1132	SSSNPSSPAPW---XKDFVTDADAEVLEHSGK	1161

RESULT 8  
A89959  
hypothetical protein SA1562 [imported] - Staphylococcus aureus (strain N315)  
C:Species: Staphylococcus aureus  
C:Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 22-Oct-2001  
C:Accession: A89959  
R:Kuroda, M.; Onta, T.; Uchivama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Ogud

ma, A.; Mizutani-Uji, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.; C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.  
Lancet 357, 1225-1240, 2001  
A:Title: Whole genome sequencing of methicillin-resistant *Staphylococcus aureus*.  
A:Reference number: A89758; MUID:213111952; PMID:11418146  
A:Accession: A89959  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-1274 <R>  
A:Cross-references: GB:A000018; EUD:gl3701536; PIDN:BA042830.1; GSPDB:GNO0149  
A:Experimental source: strain N315  
C:Genetics:  
A:Gene: SAI562

Query Match	3.2%	Score 165.5	DB 2	Length 1274
Best local similarity	20.3%	Pred. No. 0.05		
Matches	104	Conservative	75	Mismatches 173
				Indels 161
				Gaps 20
QY	49	RVPSSPA-----	CVPGGAGHR----	GSATSLVFQKTIISWMDTKGKTA 91
Db	203	RIPVSPSEKVESDKQKDYKYAKTQSNKQLEQEKQNDVVVQGTASKSSDENVSIT 262		
QY	92	ESBSLDSKNNNTRIISMSS-----	VQK-----	115
Db	263	KSMPTSKVDNTIKIENIYASQIVBEIRREKRVLQKRFFKALQKREHKNEEQDAI 322		
QY	116	-----DNFYQHNVEKLVNVSOLSKSLTEKSTOYLNQHQTAAMCKWONEKHTTEQLLESE 171		
Db	323	QRAIDEMYAKQAEYVGDSSLNDLSLTNSTDASQLHTNGI-----	E 365	
QY	172	POTVTLVPEQFSNANIDRSQNDHSDTQSEENRDNQOFLTTVKLANAKQTTBDE--	HAR 229	
Db	366	NETVS-----NDENKQASIQNEETNDTHVDESFPYVEEVLNQVSTTKQLSDDEVTSN 419		
QY	230	EAKSHQKCSKSHPGEDCASCQDEIDVVPKSPLSDVGSDDVGTGSKNDNKLIRQSCSLG 289		
Db	420	VTSQHQ-----SALQHNVENVNDKDEL-----	KNQSRLLADSEEDG 454	
QY	290	NSPPFEKESPESPMDVNDKNSCQSEADEETSPGFDEQDSSSGSSGTANKPRFQAR--	347	
Db	455	AT--NKEEYSGSQID--DAEFYELNDTEVDDETSTNIEDTNNENEMHVADPKTOEHAV 510		
QY	348	-----DADIEFRYTKTGEVHELHFQFEGGESRTGMNDLNAKLPGNISLNV 395		
Db	511	TESQVNNIDKTVNDEIELAPRHK-KDDQTNLVN-----	NSLKTNDVN----DGHV--V 556	
QY	396	ECRNSKQHGKDKSDITDHLMLPKAEDRRKEQWETKHQRTKRIKPYVPVPHLSPDK----	451	
Db	557	EDSSMNEIEKHAEIENV-----	QNEAAESE-QNVYEKTIENVNPKQTEKVSTL 606	
QY	452	-----KWLGTPIEMRRMPR---CGIRLPLLRP 476		
Db	607	SKRPENVVMTFSDKKRMWDRKGSKNVNPBLKP 639		

RESULT 9

D96730

unknown protein F5A18.20 [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001

C:Accession: D96730

R:Theologian, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso, C.; Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.; Jensen, N.F.; Hughes, B.; Huizar, L.

Nature 408, 816-820, 2000

A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.; C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali, Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.

A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, I.; Ker, M.; Wu, D.; Yu, G.; Frazer, C.M.; Venter, J.C.; Davis, R.W.

A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.

A:Reference number: A86141; PMID:21016719; PMID:11130712

A:Accession: D96730

A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-884 <STO>  
A:Cross-references: GB:AE005173; NID:96453896; PIDN:AAF09079.1; GSPDB:GN00141  
C:Genetics:  
A:Gene: F5A18.20  
A:Map position: 1

Query Match 3.2%; Score 165; DB 2; Length 884;  
Best Local Similarity 20.6%; Pred. No. 0.032;  
Matches 116; Conservative 79; Mismatches 197; Indels 170; Gaps 25;

QY 16 WGAATTS-----PAASDARSPPSRQRRV-LDPKDAHVQ-----FRVPPSPACV--PGQAG- 63  
Db 317 WQHTTSGVVPPIPPSSAQSIPOHDSMAIFPVSGHIMPPYGRFPFPNPQVGPYAFGT 376  
QY 64 -----QHRGSATSLVFKTKTITSWMDTKGKT-----AESLSL--- 97  
Db 377 KPPLHPVAAAFDDSYAASVPPKAPVFNWLBELLKKKADLGRPPSGRFRFEESMDDDV 436  
QY 98 -----SKENN-----NTRIESMMSSV-----QKDNFYQH 121  
Db 437 LYKPTKADQDKKSFSPNSDSDEEDEMADAARTTEINMEIKLITVLLKVTDLFDE 496  
QY 122 NVEKLNVNSQ-LSLDKSLTEK-----STOYLNQ-----HOTAACKWQNEKGKHTQELLESPQ 173  
Db 497 IATKVINDEAI PKDDSVQHNNHLSSLLSTADPLHKASAKILVSVEGANTKASSGSPAD 556  
QY 174 TVTLVPEQFSNANIDRSPQNDHSDTSEEN-----RDN-QOFLTTVKLANAKOTT 223  
Db 557 VLGLA-----SYASDDDDADTDAAADANADENGVSGLVGSRRHNSQDPSTKLT-----P 606  
QY 224 EDEHAREAKSHQKSKCHPCEDCASCQD-----EIDVVPKSPLSDV-----GSED 270  
Db 607 DPEAMASAKLPDPAVGVNANSKNSGLEDYSQMPGSTRKDEAGTKISDVSAASSGLDD 666  
QY 271 VGTGSKNDNKLIRQESCICLGNPPPEKESEPESPMDVNSKNSQDSEADETSPGFDEQ 330  
Db 667 DTSGSRKEH-----PDRDSDKADAILDEPHVKN-----GVKS 699  
QY 331 DGSSSQTAKNSRQARADATIEFRKX-YSTKGGEVRLHFQEGGESRTGMDNKLPGN 389  
Db 700 DCLNRQDSNKFYGDLSDEVSTDRSRIVETKGGKKGDSQ-----NDKDRMKENDLKSARK 756  
QY 390 ISSLNVECRNSKQKGDKSKITDHLMLPKAEADR-----RKEQWTKHQTERKIPKYV 443  
Db 757 VGVESNKKSTDPHVKKDSRDVERPHRTNSKEDRCRKEKEERSRHRRAENS----- 811  
QY 444 PPHLSPPDKKWLGTPIEVRMP 465  
Db 812 -----SKDKR-----RRSP 820

RESULT 10  
T41581  
hypothetical coiled-coil protein - fission yeast (Schizosaccharomyces pombe)  
C:Species: Schizosaccharomyces pombe  
C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 11-Aug-2003  
C:Accession: T41581  
R:Murphy, L.; Harris, D.; Wood, V.; Lyne, M.H.; Rajandream, M.A.; Barrell, B.G.  
submitted to the EMBL Data Library, September 1998  
A:Reference number: Z22002  
A:Accession: T41581  
A:Status: translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-4717 <MUT>  
A:Cross-references: EMBL:AL031546; PIDN:CAA20864.1; GSPDB:GN00068; SPDB:SPCC737.08  
A:Experimental source: strain 972h(-)  
C:Genetics:  
A:Gene: SPDB:SPCC737.08  
A:Map position: 3  
C:Superfamily: mldasin (AAA ATPase with von Willebrand factor type A (vwa) domain)

Query Match 3.2%; Score 164; DB 2; Length 4717;  
Best Local Similarity 19.6%; Pred. No. 0.38;  
Matches 187; Conservative 143; Mismatches 326; Indels 296; Gaps 51;

QY 90 TAESESLDSKNNNTRIESMMSSVQKDNFYQHN-----VEKLVNVSQLSLDKSL 138  
Db 3942 TANQSDLD--ESEARELESMDNGVTKDSVVSRENENSDEENQDLDDEVNDIPEDLNSL 3999  
QY 139 TEKSTOYLNOHOTAAMCKWQNEKGKHTQELLESPQTVTLVPEQFSNANIDRSPQNDH--- 196  
Db 4000 NEK-----LWDEPNE--EDLLETKS---NEQSAANNESDLVSKEDDNKA 4040  
QY 197 -STDSSENRDNOQFLTTVKLANAKOTTEDEHAREAKSHQKSKCHPCGDCASCQD-- 253  
Db 4041 LEDKROKEDEDEMSDDVGI-----DDEIQDIOENN--SQPPPENEDHLDLPDLK 4091  
QY 254 ----EIDVVPKSPLSVGSSEDDVGTGSKNDNKLIRQESCLGNSPP---FKSESEPESEMDV 306  
Db 4092 LDEKGDVSKDSLEDMDE-----ADENK---EEADAEDPEMQDFEDPLEENNTLDE 4143  
QY 307 DNSKNSCQDSEADEE--TSPGFDEQ-----EDGSSSOTA-----NKP-- 341  
Db 4144 DIQDDPSDLAEDDEKMNEDGFBNVQENBEESTEDGVKSDEELEQGEVPEQDAIDNHPKM 4203  
QY 342 ---SRFOARADATIEFRKRYSTKGGEVRLHFQF--EGGESRTGMDNKLPGNISLNV 396  
Db 4204 DAKSTFASAEADEE-----NTDKGIVGENBELGEEDGAAGSGVNG-NGTADGEFSSAEQV 4257  
QY 397 CRNSKQKQKDS-----KITDHLMLPKAEADREKQW-----ETKHQTERKIPKYV 443  
Db 4258 QKGEDTSTPKAMSEADQVQSLGDLHLEWQQA--NRHEWEDLTESQSQAFFDSEFMHV 4315  
QY 444 PPHLSPPDKKWLGT-----PIEMRMPPCGIRLPLRFSANHTVIRVDLLRAGVPEFPF 499  
Db 4316 KEDEEEDLQALGNAEKDQIKSIDRDESAN-----QNPDSMNSTNIAED--EAEVGD-- 4365  
QY 500 THYKDLWNDKHKVMPCEQNLYPVEDENGERTAGSRWELI--QTALLNKFTRPQNLKDAIL 558  
Db 4366 ---KQLQDG-----QDLSDIK-QTGEOTLTPFGSINQSEKVFELSEDEIDELP 4412  
QY 559 KYNVAYSKKWDFTALI-----DFWDKVLBEAAQHLYQSILPDMVKIALCLPNICTQPI 612  
Db 4413 DYNV---KLTNLPAAPIDEARDLWKN-HEDESTKQ-----LSIELC----- 4449  
QY 613 PLLKQKNHSITMGEOQIASILLANAFCTPPRRNAKMKSEYSSYPDINFNRLFEGRSR- 671  
Db 4450 -----EQLRLILE-----PTLATKMQGDFRTGKRLNMKKIIPYIASQF 4487  
QY 672 KPEKLTLLFCYFRVRVTEKPTGLVTFTTQSLDPEWERCCKPLTRLHVTVEGTIENGQ 731  
Db 4488 KQDKI-----WNRVKPSKRTYQVMI--SIDD-----SKMSSEGS 4521  
QY 732 GMLQVDFANRFVGGVTSAGLVQE-----EIRFLINPELI--ISRLFTEVLHDNECLIT 784  
Db 4522 TVLALETU-----ALVTKALSLEVGQIAMVKFGEQPELLHPDKQFSSE-----S 4567  
QY 785 GTEQYSEVTGVAETRYRWSRSHEDGSRDDCERRCTEIVADIALHFRYLDQFVPEKMR 844  
Db 4568 GVOMFSHFT-----FEQSNNTVLALADASKMCFN-YANTASHRNSD-----IRQ 4612  
QY 845 LNKAYCGFLRPVGSSENLSAVATGNCWCGAFGGDARLKALQILAAAAAARDVVVYFTGD 904  
Db 4613 LEI-----IISDGICEDH-----DSIRKLRRAGEEKVMIVFVILD 4648  
QY 905 SELMRDIYSMEHIFLTERKLTVDGVYKLLRYNNEECRNCSTPGPDIKIYPI 956  
Db 4649 N-----VNTQKSSILDIKV---YYDTK-----EDGTMDLKIQPIYI 4682

RESULT 11  
T29043  
hypothetical protein B0228.2 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans

[illegible]



A:Molecule type: protein  
A:Residues: 7-16;46-55;680-699;1047-1056;1921-1930 <SPI>  
R:Speicher, D.W.; Davis, G.; Marchesi, V.T.  
J. Biol. Chem. 258, 14938-14947, 1983  
A>Title: Structure of human erythrocyte spectrin. II. The sequence of the alpha-I domain  
A:Reference number: A9240B; MUID:84087888; PMID:6654896  
A:Accession: AO2965  
A:Molecule type: protein  
A:Residues: 7-601 <SPC>  
R:Speicher, D.W.; Davis, G.; Yurchenco, P.D.; Marchesi, V.T.  
J. Biol. Chem. 258, 14931-14937, 1983  
A>Title: Structure of human erythrocyte spectrin. I. Isolation of the alpha-I domain and  
A:Reference number: S1313B; MUID:84087887; PMID:6654895  
A:Accession: S1313B  
A:Molecule type: protein  
A:Residues: 7-92,'X','94'-96,'N110-151;317-342;345-366,'Z','368-370,'X','372-373;387-434;'452-'4'  
R:Lustigian, D.M.; Otaishat, N.; LaBrake, C.C.; Yu, R.N.; Davis, J.; Kelley, M.R.; Fung,  
J. Biol. Chem. 269, 25955-25958, 1994  
A>Title: The first human alpha-spectrin structural domain begins with serine.  
A:Reference number: A3892B; MUID:95014412; PMID:7929303  
A:Contents: annotation; determination of structural domain  
C:Comment: Spectrin associates with band 4.1 and actin to form the cytoskeletal superstruc-  
ture of approximately 106 residues each.

C:Genetics:  
A:Gene: GDB:SPTAL

A:Cross-references: GDB:119601; OMIM:182860

A:Map position: lq21-lq21

C:Superfamily: spectrin alpha chain; calmodulin repeat homology; SH3 homology; spectrin/  
F:52-156/Domain: spectrin/dystrophin repeat homology <SPI>  
F:157-262/Domain: spectrin/dystrophin repeat homology <SP1>  
F:263-368/Domain: spectrin/dystrophin repeat homology <SP3>  
F:369-474/Domain: spectrin/dystrophin repeat homology <SP4>  
F:475-580/Domain: spectrin/dystrophin repeat homology <SP5>  
F:581-685/Domain: spectrin/dystrophin repeat homology <SP6>  
F:686-791/Domain: spectrin/dystrophin repeat homology <SP7>  
F:792-897/Domain: spectrin/dystrophin repeat homology <SP8>  
F:898-983/Domain: spectrin/dystrophin repeat homology #status atypical <SP9>  
F:984-1031/Domain: SH3 homology <SH3>

F:1081-1181/Domain: spectrin/dystrophin repeat homology <SP10>  
F:1182-1287/Domain: spectrin/dystrophin repeat homology <SP11>  
F:1288-1393/Domain: spectrin/dystrophin repeat homology <SP12>  
F:1394-1498/Domain: spectrin/dystrophin repeat homology <SP13>  
F:1499-1604/Domain: spectrin/dystrophin repeat homology <SP14>  
F:1605-1710/Domain: spectrin/dystrophin repeat homology <SP15>  
F:1711-1816/Domain: spectrin/dystrophin repeat homology <SP16>  
F:1817-1925/Domain: spectrin/dystrophin repeat homology <SP17>  
F:1926-2032/Domain: spectrin/dystrophin repeat homology <SP18>  
F:2041-2146/Domain: spectrin/dystrophin repeat homology <SP19>  
F:2155-2257/Domain: spectrin/dystrophin repeat homology <SP20>  
F:2270-2302/Domain: spectrin/dystrophin repeat homology <EF1>  
F:2313-2345/Domain: calmodulin repeat homology <EF2>



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 26, 2004, 17:20:49 ; Search time 15.6372 Seconds  
(without alignments)

3249.964 Million cell updates/sec

Title: US-09-302-812-4

Perfect score: 5190

Sequence: 1 MNAGGCEPCTKATRWGAAT.....YHVESCAETADHSQRTGT 976

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_42.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	178	3.4	1359	1 ATRX CAEEL	Q9y7e0 caenorhabdi
2	174	3.4	5147	1 PCLO HUMAN	Q9y6v0 homo sapien
3	170.5	3.3	5038	1 PCLO_MOUSE	Q9qyx7 mus musculu
4	168	3.2	5085	1 PCLO_RAT	Q9yjs6 ratus norv
5	162	3.1	2492	1 ATRX PANTR	Q7yqm4 pan troglod
6	161	3.1	2476	1 ATRX_MOUSE	Q61687 mus musculu
7	161	3.1	2492	1 ATRX_HUMAN	P46100 homo sapien
8	161	3.1	2492	1 ATRX_PONPY	Q7ygm3 pongo pygma
9	159.5	3.1	1183	1 BMS1_YEAST	Q08965 saccharomyc
10	159	3.1	406	1 SR40_YEAST	P32583 saccharomyc
11	159	3.1	1462	1 NKCR_HUMAN	P30414 homo sapien
12	159	3.1	2418	1 SPCA_HUMAN	P02549 homo sapien
13	157.5	3.0	1311	1 ATRX_DROME	Q9gqns drosophila
14	155	3.0	1138	1 BM2K_MOUSE	Q91296 mus musculu
15	154.5	3.0	646	1 SGL_BOVIN	P23389 bos taurus
16	154.5	3.0	3259	1 GOB1_HUMAN	Q14789 homo sapien
17	154	3.0	1972	1 MYHB_MOUSE	O08638 mus musculu
18	153.5	3.0	1878	1 BRCL_CANFA	Q95153 canis faml1
19	153	2.9	1031	1 YDG9_SCHPO	Q10496 schizosacch
20	153	2.9	6632	1 UN89_CAEEL	O01761 caenorhabdi
21	152	2.9	1447	1 BUDA_YEAST	P47136 saccharomyc
22	151	2.9	2415	1 SPCA_MOUSE	P08032 mus musculu
23	150.5	2.9	1312	1 RA50_YEAST	P12753 saccharomyc
24	150.5	2.9	1957	1 SPOF_SCHPO	Q10411 schizosacch
25	150.5	2.9	1972	1 MYHB_RABIT	P35748 oryctolagus
26	150	2.9	2137	1 SPCB_HUMAN	P11277 homo sapien
27	150	2.9	3969	1 HRX_HUMAN	Q03164 homo sapien
28	149.5	2.9	1210	1 AF4_HUMAN	R51825 homo sapien
29	148.5	2.9	472	1 YWIE_CAEEL	Q23525 caenorhabdi
30	148.5	2.9	677	1 SGL_HUMAN	P05060 homo sapien
31	148.5	2.9	1928	1 MYS1_YEAST	P08964 saccharomyc
32	148	2.9	644	1 NFM_RABIT	P54938 oryctolagus
33	148	2.9	1679	1 GCC3_MOUSE	Q08chg3 mus musculu

## ALIGNMENTS

### RESULT 1

ID	ATRX CAEEL	STANDARD;	PRT;	1359 AA.
AC	Q9U7E0; O02061;			
DT	28-FEB-2003 (Rel. 41, Created)			
DT	28-FEB-2003 (Rel. 41, Last sequence update)			
DT	15-MAR-2004 (Rel. 43, Last annotation update)			
DE	Transcriptional regulator ATRX homolog (X-linked nuclear protein-1).			
GN	XNP-1 OR B0041.7.			
OS	Caenorhabditis elegans.			
OC	Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;			
OC	Rhabditidae; Peloderinae; Caenorhabditis.			
OX	NCBI_TaxID=6239;			
RN	[1]			
RP	SEQUENCE FROM N.A., AND CHARACTERIZATION.			
RX	MEDLINE=99365296; PubMed=10433961;			
RA	Villard L., Fontes M., Ewbank J.J.;			
RT	"Characterization of xnp-1, a Caenorhabditis elegans gene similar to the human XNP/ATR-X gene.";			
RL	Gene 236:13-19(1999).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=Bristol N2;			
RA	Fulton R., Wohldmann P.;			
RL	Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.			
CC	- - FUNCTION: Could be a global transcriptional regulator. Modifies			
CC	gene expression by affecting chromatin (Potential).			
CC	- - SUBCELLULAR LOCATION: Nuclear.			
CC	- - SIMILARITY: Belongs to the SNF2/RAD54 helicase family.			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).			
CC	EMBL; AF134186; RAD55361.1; -;			
CC	EMBL; AF000196; RAC24256.1; -;			
DR	PIR; T34036; T34036.			
DR	WormPep; B0041.7; CEI7314.			
DR	InterPro; IPR001410; DEAD.			
DR	InterPro; IPR001650; Helicase_C.			
DR	InterPro; IPR000330; SNF2_N.			
DR	Pfam; PF00271; helicase_C; 1.			
DR	Pfam; PF00176; SNF2_N; 1.			
DR	SMART; SM00487; DEXDC; 1.			
DR	SMART; SM00490; HELIC; 1.			
DR	PROSITE; PS00690; DEAD_ATP_HELICASE; FALSE NEG.			
KW	DNA repair; Hydrolase; Nuclear protein; ATP-binding;			
KW	DNA-binding; 496 503			
FT	NP_BIND	636 639		
FT	DEAH_BOX	67 70		
FT	POLY-ASP.	266 272		
FT	POLY-GLU.			

34	146.5	2.8	667	1	CYLL_BOVIN	P35662 bos taurus
35	146.5	2.8	3214	1	EPAL_HUMAN	O03001 homo sapien
36	146	2.8	5120	1	PCLO_CHICK	Q9pu36 gallus gall
37	144.5	2.8	1980	1	MY9B_RAT	Q63358 ratus norv
38	144.5	2.8	2319	1	AKA6_HUMAN	Q13023 homo sapien
39	144	2.8	1637	1	MRSP_STAAU	P80544 staphylococ
40	143.5	2.8	743	1	ABRA_PLAFC	P22620 plasmodium
41	143.5	2.8	1338	1	ACIN_MOUSE	O9jix8 mus musculu
42	143.5	2.8	1969	1	MYSA_CAEEL	P12844 caenorhabdi
43	143.5	2.8	4835	1	MONI_GIALA	Q8t5t1 giardia lam
44	143	2.8	1085	1	YAPA_SCHPO	Q09863 schizosacch
45	143	2.8	1433	1	REBT_CHICK	O42184 gallus gall

```

FT DOMAIN 276 281 POLY-LYS.
FT DOMAIN 372 375 POLY-LYS.
FT DOMAIN 603 608 POLY-LYS.
FT DOMAIN 859 862 POLY-LYS.
FT CONFLICT 479 479 C -> F (IN REF. 2).
SQ SEQUENCE 1359 AA; 156191 MW; EB4342547D4F4E64 CRC64;

Query Match 3.4%; Score 178; DB 1; Length 1359;
Best Local Similarity 19.1%; Pred. No. 0.01;
Matches 147; Conservative 118; Mismatches 279; Indels 224; Gaps 33;

QY 86 KGITAESELSKNNNTPIESMSVQNDYQHNVEKLVNVSQISDKSLTEKSTQY 145
Db 81 KSRKRAKSES-ESDESDEEDRKSKKKVDQKKEKSKKKTTSSEDEDEEREQ- 138
QY 146 LNQHQAAMCKQNEGHTEQLLESEPTVTLVPEQFSNANIDRSQNDHSDTDSEENR 205
Db 139 -KSKKSKTKTKOTSSSESE---ESEERKVKKSKKKEKSKVKRAETSESEDEKPSK 194
QY 206 DNOQFTTIVKLANAKOTTDEHAREAKSHOKSKCHPGEDCASCQODEIDVVPKPSLSD 265
Db 195 KSKKGLK--KKAKESESESEDEKVKSKKKSKKVVKKSESEDEDEAPEKKTERRSK 252
QY 266 VGSDEVGTGKNNKLIQROSCLGNSPPPEKESEP-----ESPMDVNSKNSCOOSEADEE 321
Db 253 TSSESESESEKSDDE--ER-----EKSSPKPKKKPLAVKKLSDESESESDVE 300
QY 322 TSP-----GFDQEDGSSSOTANKPSRFOARDADIEFRKRYSTKGGEVRLHF 368
Db 301 VLPQKKRGAVTILSDEDEKQKSESEADVEKYSKKA-----KKQESSE----- 348
QY 369 QFEGGSRGTGMDNLAKLPNGISLNVCEKNSKHG--KKDSKITDMLMLPKAEDRRKE 426
Db 349 --SGSDSESGSITVNRK-----SKKKEPEKKKKGIIMDSKLOKETIDAERAERK 400
QY 427 QWETKHQR-----TERKPKVYPHPSLSPDKKWLGPPIEMRMMPR 466
Db 401 RLEKKQKEFGNIVLEGEDITLMTGTSSQRLKSVV---LDPE-----SSTVDSESKP- 452
QY 467 CGIRLPLRPSANHTVIRVILLRAGVRPKPFPPTHYKDLMDNKHVMKPCSEQNLXPVEDE 526
Db 453 -----VEVHNSLVR--ILKPHQAH-----GIQPMYDCACE 480
QY 527 NGER--TAGSRWELIOTALLNKFTPONLKDAILKNVAYKKWDFALIDFMDKXVLEEA 584
Db 481 SLDRLEDEGS-----GGILAHCMGLGKT---LQVITFLHTVL--- 514
QY 585 EAQHLIQSILPDMVKIALCLP-NICTQPIPLKKNNHSITMSQEQIATSLANAFFCTFP 643
Db 515 -----MHKEIGYCKKRVLVVVPKNVINWF-----KEFOKWLVDNEEDLTIDVN----- 559
QY 644 RNAKMKSESYSPDINFNR--LPEGRSSRKPEKLTLCFYFRVFT-----EKKPTGLV 695
Db 560 -----ELDSYKTIEDRRALKAWHSSKTPSVMIIGYDLFRILTVEDDPKPKKPNRN 611
QY 696 TFRQSLDF-----PWERCE-----KPLTLHTYEGTIEHG 730
Db 612 RLLEKAKEDFRKYLQNPDPMDVVCDEAHKLNKDSALSCKMVKILTKRRLCTGTPLQNN 671
QY 731 -----QGMVL--QVDFANRFVGGVTSAGLVQE---EIREL 760
Db 672 LMEYHVMVNFVKPGLLGTGKTEFANRFV--NIINRGRTKDSAPLEVSFM 717

```

RESULT 2

```

PCLO_HUMAN STANDARD; PRT; 5147 AA.
AC Q9Y6V0; O43373; O60305; Q9BVC8; Q9UIV2; Q9Y6U9;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Piccolo protein (Aczonin) (Fragment8).
GN PCLO OR ACZ OR KIAA0559.

OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]_TaxID=9606;
RP SEQUENCE OF 1-759 FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=99439764; PubMed=10508862;
RA Wang X., Kibschull M., Laue M.M., Lichte B., Petrasch-Parwez E.,
RA Killmann M.W.;
RT "Aczonin, a 550-kd putative scaffolding protein of presynaptic active
RT zones, shares homology regions with rim and bassoon and binds
RT profilin.";
RL J. Cell Biol. 147:151-162(1999).
RN [2]
RP SEQUENCE OF 552-4404 FROM N.A.
RA Kraemer J., Wollam C., Wohldmann P., McGrane B.;
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 3619-5147 FROM N.A. (ISOFORM 2).
RC TISSUE=Brain;
RX MEDLINE=98290545; PubMed=9628581;
RA Nagase T., Ishikawa K.-I., Miyajima N., Tanaka A., Kotani H.,
RA Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. IX.
RT The complete sequences of 100 new cDNA clones from brain which can
RT code for large proteins in vitro.";
RL DNA Res. 5:31-39(1998).
RN [4]
RP SEQUENCE OF 4405-4439 FROM N.A.
RC TISSUE=Placenta;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Scapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Udwin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodargren E.J., Lu X., Gibbs R.A.,
RA Faney J., Heiton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
RT human and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [5]
RP SEQUENCE OF 4405-5147 FROM N.A.
RA Kalicki J., Elliott G.;
RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: May act as a scaffolding protein involved in the
CC organization of synaptic active zones and in synaptic vesicle
CC trafficking (By similarity).
CC -!- SUBUNIT: Interacts with Rabac1/Pral, RIMS2 and profilin (By
CC similarity).
CC -!- SUBCELLULAR LOCATION: Concentrated at the presynaptic side of
CC synaptic junctions (By similarity).
CC -!- ALTERNATIVE PRODUCTS:
CC Event-Alternative splicing; Named isoforms=2;
CC Comment=Additional isoforms seem to exist;
CC Name=1;
CC IsoId=Q9Y6V0-1; Sequence=Displayed;
CC Name=2;
CC IsoId=Q9Y6V0-2; Sequence=VSP_003923, VSP_003924, VSP_003925,
CC VSP_003926, VSP_003927;
CC Note=No experimental confirmation available;
CC -!- DOMAIN: C2 domain 1 is involved in binding calcium and

```



phospholipids. Calcium binds with low affinity but with high specificity and induces a large conformational change.

-1- SIMILARITY: Contains 2 C2 domains.

-1- SIMILARITY: Contains 1 PDZ/DRH domain.

THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.ebi.ac.uk/announcements> or send an email to [license@ebi.ac.uk](mailto:license@ebi.ac.uk)).

EMBL; Y13188; CAB60727.1; --  
EMBL; AC004903; AAD20936.1; --  
EMBL; AC004886; AAD21789.1; --  
EMBL; AB011131; BAA25485.1; --  
EMBL; BC001304; AAH01304.1; --  
EMBL; AC004082; AAB97937.1; --  
PIR; T00634; T00634.  
HSP; P04410; I225.  
Genew; HGNC:13406; PCLO.  
MIN; 604918; --  
GO; GO:0005856; C:cytoskeleton; NAS.  
GO; GO:0045202; C:synaptic junction; ISS.  
GO; GO:0005509; F:calcium ion binding; ISS.  
GO; GO:0005544; F:calcium-dependent phospholipid binding; ISS.  
GO; GO:0005522; F:profilin binding; ISS.  
GO; GO:0007010; P:cytoskeleton organization and biogenesis; ISS.  
GO; GO:0016080; P:synaptic vesicle targeting; ISS.  
InterPro; IPR000008; C2.  
InterPro; IPR001565; Synaptotagmin.  
PRINTS; PR00360; C2DOMAIN.  
PRINTS; PR00399; SYNAPTOTAGMIN.  
SMART; SM00239; C2; 2.  
PROSITE; PS00499; C2 DOMAIN 1; 1.  
PROSITE; PS00004; C2 DOMAIN 2; 2.  
Calcium/phospholipid-binding; Zinc; Metal-binding; Zinc-finger;  
Repeat; Alternative splicing.

NON TER 1 1  
DOMAIN 400 465

10 X 10 AA TANDEM APPROXIMATE REPEATS OF  
P-A-K-P-Q-P-Q-P-X.  
C4-TYPE (POTENTIAL).  
C4-TYPE (POTENTIAL).

POLY-PRO.  
PDZ.  
C2 DOMAIN 1.  
C2 DOMAIN 2.

S -> SGNGLGIRIVGKEIPHSGEIGAYIAKILPGSSAE  
/FTID-VSP\_003923.  
/FTID-VSP\_003923.

K -> KPTDGTGVVSHPTGIEQ (in isoform 2).  
/FTID-VSP\_003924.  
G -> GQVMVQNAS (in isoform 2).  
/FTID-VSP\_003925.

TAHKS -> SKRKK (in isoform 2).  
/FTID-VSP\_003926.  
Missing (in isoform 2).  
/FTID-VSP\_003927.

SEQUENCE 5147 AA; 563537 MW; CD584990498CD3C CRC64;

Query Match 3.4%; Score 174; DB 1; Length 5147;

Best Local Similarity 19.7%; Pred. No. 0.099;

Matches 140; Conservative 120; Mismatches 251; Indels 198; Gaps 32;

9 PCTATRWGAATTSAAADSFQRORVLDPKDAHQVFRVPPSPACVPGQAQHRGS 68

1227 PCT-----AKPQEKEDKSDTSSSQ-----PKSPQLSDTGYSSDGI 1265

69 ATSLVFKQKTTSMWMT-----KGIK--TAESLSLDSKENNTRIESMMSSVQKDNFYQ 120

1266 SSSL-----GEIPSLIPTDEKILKGLKXDSFQSSPSSPSDLAKLESTVLSILE----- 1316

QY 121 HNVKLVNVSQSLDLSLTKSTOYLNOHQAAMCKWQNEKGHTLEQLESEPTQVTLVPE 180  
DB 1317 -----AQASTLADEKS--EKTKQ---PHEVSP--EQPKQOEKQSL--SETLEITISEE 1361  
QY 181 QFSNANIDR-----SPQNDHSDTDSENRDNOQFLTTVK--LANAKQTTEDSHA----- 228  
DB 1362 EIKESQERKDTFKKDSQDIPSSKDHKESEFVDDITTRREPYSVESSESENPSVPQ 1421  
QY 229 REAKSHQCKSKSPGDCASCQODEIDVVPKSPLSVSGSEVDVGTGSKNDKILRQESCL 288  
DB 1422 RKRTYSVSSSSSEYKQEDSQSGSEEDFIRKQ--IIEMSADEGASGSEDDFIRNQLKEI 1480  
QY 289 GNSPPEKSEPE-----SPMDVNSKNSCODSEADETSPGDFDQE 330  
DB 1481 SSSTESOKKETGKGIKTAGKHRLTRKSTSTDEAGRRHSHWDEDE-----AFDESP 1536  
QY 331 DGSSSQTKANKPS-----RFOARDADIEFRKRYSTKGGEVRLHFOF----- 370  
DB 1537 ELKYRETKSQESEELVVTGGGLRRFKTIELNSTIADKYSAESSQKKTSLYFDEEPELEM 1596  
QY 371 ----EGGESRT---GMNDLNK-----LPCNISILNVECHNSKHGKDKSK-----IT 411  
DB 1597 ESLTDSPEDRSGEGSSLSHASTFTPTSTSVSLDESDSPSHKKGESKQORKAKHR 1656  
QY 412 DHLMLPKAEDR-----RKEOWETKHQTERKI----- 439  
DB 1657 PHGLPLLTIEDSSEELREBELLEKQEKQRELEQOQRKSSSKSKKDKDELRAQRRE 1716  
QY 440 -PKYVPHLSPDKKWLGTPIEEMRMPCGIRLPLLRPS--ANHTVTIRVDILLRAGEVPKP 497  
DB 1717 RPKTPPSNLSPID--ASPTLELRQAAME---ELHRSSCSEYSPSIESD-----PEG 1764  
QY 498 PPTHYKOLMD-NKHVMPKCSNLYPVEDENGERTAGRWELIQTALLNKETRPQNLKDA 556  
DB 1765 FEISPEKIIIEVQYKLP-TAVSLYSPTDEQ-----SIMQKEGSKALKASA 1809  
QY 557 ILKYNVAYSKKWDFAL--IDFWDKVLEAE-----AQHLYQSILPD 596  
DB 1810 EEMTEEMMHKTHKYKAPPAANERDEVEFEKPLYGGMLEIDYIYSLVED 1858

## RESULT 3

PCLO MOUSE STANDARD; PRF; 5038 AA.  
AC Q9QYX7; Q9QYX6; Q9QZJ0;  
DT 28-FEB-2003 (Rel. 41, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Piccolo protein (Presynaptic cytomatrix protein) (Aczonin) (Brain-derived HLMN protein).  
DE PCLO OR ACZ.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
[1]  
SEQUENCE FROM N.A., SUBCELLULAR LOCATION, ALTERNATIVE SPLICING,  
TISSUE SPECIFICITY, AND INTERACTION WITH PROFILIN.  
RC TISSUE=Brain;  
RX MEDLINE=99439764; PubMed=10508862;  
RA Wang X., Kibschull M., Laue M.M., Lichte B., Petrasch-Parwez E., Kilimann M.W.;  
RA "Aczonin, a 550-kd putative scaffolding protein of presynaptic active zones, shares homology regions with rim and bassoon and binds profilin."  
RL J. Cell Biol. 147:151-162(1999).  
[2]  
REVIEWS.  
RA Kilimann M.W.;  
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.  
[3]  
RP SEQUENCE OF 4502-4692 FROM N.A.

TISSUE=Brain;  
Huang W., Jin W., Huang C., Chen B., Zhang J., Ju G.;  
Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.  
[4]  
RN INTERACTION WITH RIMS2.  
RA MEDLINE=22384373; PubMed=12401793;  
RX Fujimoto K., Shibasaki T., Yokoi N., Kashima Y., Matsumoto M.,  
RA Sasaki T., Tajima N., Iwanaga T., Seino S.;  
RT "Piccolo, a Ca<sup>2+</sup> sensor in pancreatic beta-cells. Involvement of  
RT CAMP-GEFII, RimM, Piccolo complex in CAMP-dependent exocytosis.";  
RL J. Biol. Chem. 277:50497-50502(2002).  
CC -!- FUNCTION: May act as a scaffolding protein involved in the  
CC organization of synaptic active zones and in synaptic vesicle  
CC trafficking.  
CC -!- SUBUNIT: Interacts with Rabac1/Pral, RIMS2 and profilin.  
CC -!- SUBCELLULAR LOCATION: Concentrated at the presynaptic side of  
CC synaptic junctions.  
CC -!- ALTERNATIVE PRODUCTS:  
CC Event=Alternative splicing; Named isoforms=2;  
CC Name=1;  
CC IsoId=Q9QYX7-1; Sequence=Displayed;  
CC Name=2;  
CC IsoId=Q9QYX7-2; Sequence=VSP\_003928, VSP\_003929;  
CC -!- TISSUE SPECIFICITY: Highly expressed in brain. Low levels found in  
CC stomach. Not detected in other tissues analyzed including adrenal  
CC gland, testis and pancreas.  
CC -!- DOMAIN: C2 domain 1 is involved in binding calcium and  
CC phospholipids. Calcium binds with low affinity but with high  
CC specificity and induces a large conformational change.  
CC -!- SIMILARITY: Contains 2 C2 domains.  
CC -!- SIMILARITY: Contains 1 PDZ/DHR domain.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC -----  
CC EMBL; Y19185; CAB60731.2; --  
CC EMBL; Y19186; CAB60732.2; --  
CC EMBL; AF181269; AAD55786.2; --  
CC HSSP; P04410; IA25.  
CC MGD; MGI:1349390; Pclo.  
CC GO; GO:0045202; C:synaptic junction; IDA.  
CC GO; GO:0005509; F:calcium ion binding; ISS.  
CC GO; GO:0005544; F:calcium-dependent phospholipid binding; ISS.  
CC GO; GO:0005522; F:profilin binding; IDA.  
CC GO; GO:0019933; P:CAMP-mediated signaling; IDA.  
CC GO; GO:0007010; P:cytoskeleton organization and biogenesis; IDA.  
CC GO; GO:0030073; P:insulin secretion; IDA.  
CC GO; GO:0017157; P:regulation of exocytosis; IDA.  
CC GO; GO:0016080; P:synaptic vesicle targeting; NAS.  
CC InterPro; IPR000008; C2.  
CC InterPro; IPR001478; PDZ.  
CC InterPro; IPR008899; Znf\_piccolo.  
CC Pfam; PF00168; C2; 2.  
CC Pfam; PF00595; PDZ; 1.  
CC Pfam; PF05715; Zf\_piccolo; 2.  
CC SMART; SM00239; C2; 2.  
CC SMART; SM00228; PDZ; 1.  
CC PROSITE; PS00499; C2 DOMAIN 1; 1.  
CC PROSITE; PS00004; C2 DOMAIN 2; 2.  
CC PROSITE; PS0106; PDZ; 1.  
CC Calcium/phospholipid-binding; Metal-binding; Zinc; Zinc-finger;  
CC Repeat; Alternative splicing.  
CC  
CC 10 X 10 AA TANDEM APPROXIMATE REPEATS OF  
CC DOMAIN 371 470  
CC P-A-K-P-Q-P-Q-P-X.  
CC C4-TYPE (POTENTIAL).  
CC C4-TYPE (POTENTIAL).  
CC POLY-PRO.  
CC PDZ.  
CC  
CC ZN\_FING 502 526  
CC ZN\_FING 967 990  
CC DOMAIN 2305 2329  
CC DOMAIN 4394 4488

FT DOMAIN 4607 4705 C2 DOMAIN 1.  
FT DOMAIN 4922 5012 C2 DOMAIN 2.  
FT VARSPLIC 4829 4833 TKPTN -> SKRRK (in isoform 2).  
FT /FTID=VSP\_003928.  
FT VARSPLIC 4834 5038 Missing (in isoform 2).  
FT /FTID=VSP\_003929.  
SQ SEQUENCE 5038 AA; 547600 MW; DADA460CF3B40888 CRC64;  
Query Match 3.3%; Score 170.5; DB 1; Length 5038;  
Best Local Similarity 19.6%; Pred. No. 0.15;  
Matches 141; Conservative 122; Mismatches 271; Indels 187; Gaps 33;  
QY 5 PCGEPT-----KATRWGAATSP-----AASDARSFPRQRVLDPKDAHVQFRVPPSS 54  
DB PGATPTQLPKRQKESR---DVTQPAEGTAKRGEKPSKDRTEKEEDSDTSSSQPKS 1262  
QY 55 PACVPQAGQHRGSAATSLVFKQKTITSMMD---TKGIK---TAESSELDKNNNTNRI 108  
DB 1263 PQGLSDTGVSSDGIISGLISLIPS--DEKLLKGLKDSFSQESSFSSPDLAKLES 1320  
QY 109 MMSSV---QKDNFYQHNVEKLVNVSQSLDKLSTEKSTOYLNOHQHTAAMCKWQNEKHT 165  
DB 1321 TVLSILEAQASTLVGEKAEKKTQPKVSPQDQOKTQTPSETRDISIE-----E 1372  
QY 166 QLLESEPOVTVLVPQFNSNANIDRSPONDHDSDTSEENRDNQOFLTITVKLANA----K 220  
DB 1373 EIKESQKKVT-----SKDSAQGFPSRKEHKEPDELVDLSPRRASVDSVE 1419  
QY 221 QTTEDEHAREAKSHOK---CSKSHCPGEDCASCQDEIDVVPKSPSLDSVGESEVGTGS 275  
DB 1420 DSESENSPVARKKRTISGSSSEYKQEDSGSGEDE-DFIRKQ-IIEMSADEASGS 1477  
QY 276 KDNKLIRQESCLNPPPEKESEP-----ESPMQ-----VDNSKNSCQDS-----EA 318  
DB 1478 EDEEFIRSQLKEIGGVTEQKRETKGKSPAGKHRLTRKSTSTFDDDAGRHHSHWIDE 1537  
QY 319 DEETSGFDEQEDGSSSQANKPS-----RQARDADIEFRKRYSTKGEVEL 366  
DB 1538 DDET---FDESELKPRETKSQESELVAVGGGLRRFKTIELNSTVTDKYSAESSQKT 1594  
QY 367 HFOF-----EGGESRT---GMNDLNAK-----LPGNISLNVCEKNSKHGX 405  
DB 1595 TLYFDEPELEMSLTDSPEDSRGSGSSLSHASTPCTSPTSVSLDESDSSPSHKK 1654  
QY 406 KQSK-----ITDHLRLPKAEDR-----RKQWETKHQTERKI-----439  
DB 1655 GESKQQRKARHRSHGPLLPTIEDSSEEEELREBEELLKEQKQRELEQQQKSSSKSKX 1714  
QY 440 -----PKYVPHLSPKKMLGTPIEEMRRMPRCGIRLPLLRPS-ANHTVTIR 485  
DB 1715 DKDELRAQRERRRPKTPPSNLSPIED--ASPTTELQAAEME---ELHRSSCSEYSPSIE 1769  
QY 486 VDLLRAGEVPKPPFTHYKDLWD-NKHYQMPSCQNLYPVEDENGERTAGSRWELIQTALL 544  
DB 1770 SD-----PEGFEISPEKIIIEVKVYKLP-TAVSLYSPDTEQ-----SVM 1807  
QY 545 NKTRPQNKLKALKNNVAYSKKWDFTAL--IDFNKYLEEAE-----AQHLYQSILP 595  
DB 1808 QKEGAQKALKSAEMEYEMMHKPKYKAFPAANERDEVEFEKFLYGGMLIEDYIESLVE 1867  
QY 596 D 596  
DB 1868 D 1868  
RESULT 4  
PCLO\_RAT  
ID PCLO\_RAT STANDARD; PRT; 5085 AA.  
AC Q9UKS6; Q9JUL1;  
DT 28-FEB-2003 (Rel. 41, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Piccolo protein (Multidomain presynaptic cytomatrix protein).

GN PCLO.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
OX NCBI\_TaxID=10116;  
RN [1]  
RP SEQUENCE FROM N.A. (ISOFORM 2), AND INTERACTION WITH RABAC1.  
RX MEDLINE=20170257; PubMed=10707984;  
RA Fenster S.D., Chung W.J., Zhai R., Cases-Langhoff C., Voss B.,  
RA Garner A.M., Kaempf U., Kindler S., Gundelfinger E.D., Garner C.C.;  
RT "Piccolo, a presynaptic zinc finger protein structurally related to  
RT bassoon,"  
RL Neuron 25:203-214 (2000).  
RN [2]  
RP SEQUENCE FROM N.A. (ISOFORM 1).  
RA Fenster S.D., Cases-Langhoff C., Gundelfinger E.D., Garner C.C.;  
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP CALCIUM-BINDING ACTIVITY, AND MUTAGENESIS OF ASP-4668; ASP-4674;  
RP VAL-4688; MET-4689; VAL-4690; SER-4691; GLN-4692; ASN-4693 AND  
RP ALA-4694.  
RX MEDLINE=21181819; PubMed=11285225;  
RA Gerber S.H., Garcia J., Rizo J., Suedhof T.C.;  
RT "An unusual C(2)-domain in the active-zone protein piccolo:  
RT implications for Ca(2+) regulation of neurotransmitter release.";  
RL EMBO J. 20:1605-1619 (2001).  
CC -!- FUNCTION: May act as a scaffolding protein involved in the  
CC organization of synaptic active zones and in synaptic vesicle  
CC trafficking (By similarity).  
CC -!- SUBUNIT: Interacts with Rabac1/Pral, RIMS2 and profilin (By  
CC similarity).  
CC -!- SUBCELLULAR LOCATION: Concentrated at presynaptic side of synaptic  
CC junctions.  
CC -!- ALTERNATIVE PRODUCTS:  
CC Event=Alternative splicing; Named isoforms=2;  
CC Name=1;  
CC IsoId=Q9UKS6-1; Sequence=Displayed;  
CC Name=2;  
CC IsoId=Q9UKS6-2; Sequence=VSP\_003930, VSP\_003931;  
CC -!- DOMAIN: C2 domain 1 is involved in binding calcium and  
CC phospholipids. Calcium binds with low affinity but with high  
CC specificity and induces a large conformational change.  
CC -!- SIMILARITY: Contains 2 C2 domains.  
CC -!- SIMILARITY: Contains 1 PDZ/DHR domain.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement. (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; AF138789; AAF07822.2; -;  
CC EMBL; AF227534; AAF63196.1; -;  
CC HSP; P04410; 1A25.  
CC GO; GO:0045202; C:synaptic junction; IDA.  
CC GO; GO:0005509; F:calcium ion binding; IDA.  
CC GO; GO:0005544; F:calcium-dependent phospholipid binding; IDA.  
CC GO; GO:0005522; F:profilin binding; ISS.  
CC GO; GO:0007010; P:cytoskeleton organization and biogenesis; ISS.  
CC GO; GO:0016080; P:synaptic vesicle targeting; NAS.  
CC InterPro; IPR000008; C2.  
CC InterPro; IPR001478; PDZ.  
CC InterPro; IPR008899; Znf\_piccolo.  
CC Pfam; PF00168; C2; 2.  
CC Pfam; PF00595; PDZ; 1.  
CC Pfam; PF05715; Zf\_piccolo; 2.  
CC SMART; SM00239; C2; 2.  
CC SMART; SM00228; PDZ; 1.  
CC PROSITE; PS00499; C2\_DOMAIN\_1; 1.  
CC PROSITE; PS00004; C2\_DOMAIN\_2; 2.  
CC PROSITE; PS0106; PDZ; 1.

KW Calcium/phospholipid-binding; Metal-binding; Zinc; Zinc-finger;  
KW Repeat; Alternative splicing.  
FT DOMAIN 372 491  
FT 12 X 10 AA TANDEM APPROXIMATE REPEATS OF  
FT P-A-X-P-Q-P-Q-P-X.  
FT C4-TYPE (POTENTIAL).  
FT C4-TYPE (POTENTIAL).  
FT POLY-PRO.  
FT PDZ.  
FT C2 DOMAIN 1.  
FT C2 DOMAIN 2.  
FT TKPTN -> SKRKK (in isoform 2).  
FT /FTID-VSP\_003930.  
FT Missing (in isoform 2).  
FT /FTID-VSP\_003931.  
FT D->A: COMPLETE LOSS OF CALCIUM-BINDING  
FT AND CALCIUM-DEPENDENT PHOSPHOLIPID  
FT BINDING ACTIVITY.  
FT D->A: COMPLETE LOSS OF CALCIUM-BINDING  
FT AND CALCIUM-DEPENDENT PHOSPHOLIPID  
FT BINDING ACTIVITY.  
FT V->S: SMALL INCREASE IN AFFINITY FOR  
FT CALCIUM.  
FT VM->SS: 10-FOLD INCREASE IN AFFINITY FOR  
FT CALCIUM.  
FT M->S: INCREASED AFFINITY FOR CALCIUM.  
FT VV->SS: 10-FOLD INCREASE IN AFFINITY FOR  
FT CALCIUM.  
FT QN->AA: MODERATE INCREASE IN AFFINITY FOR  
FT CALCIUM.  
FT A->S: NO EFFECT ON CALCIUM-BINDING  
FT ACTIVITY.  
FT SQ SEQUENCE 5085 AA; 552702 MW; 5A1BB543201A7450 CRC64;  
Query Match 3.2%; Score 168; DB 1; Length 5085;  
Best Local Similarity 19.6%; Pred. No. 0.22;  
Matches 142; Conservative 117; Mismatches 272; Indels 192; Gaps 34;  
QY 5 PCCEPT--KATRWGAATTSPAA---SDARSPSRQRRVLDPKDAHVQFRVPPSPACV 58  
DB 1249 PQTATPLTLPKAGEKERAQAQPAQAGSSKDGQGERSEKTEKEDKSDTSSQPKSPQGL 1308  
QY 59 PQQAQGHRSATSLVFKQKTTITSMW---TKGKIK--TAESRLDSKENNNTRIEMMS 112  
DB 1309 SDTGSSYSDGSGSLGEIPLSLPS--DEKLLKGLKKDSFQSPPSPSLAKLESTVLS 1366  
QY 113 V---QKDNFYQHNVEKLVNVSQSLDKSLTEKSTQVYNQHOTAAKCKWQEGKHTQLE 169  
DB 1367 ILEAQASTLVGKAEKKTQPKISPEKPDQKQKTQTASETLDTITSE-----EIKK 1418  
QY 170 SEPTQTVTLVPEQFSNANI-----DRSPQNDHSDTDSSENRDQOQFLTVYK 215  
DB 1419 SEKKVS--PKDSQGGPPSRKEKHEKPELVDDLSPRRASYDSVEDSESENSESPVVRKK 1476  
QY 216 LANAQTTDEHARBAKSHQCKSPCHPGEDCASQOQDIDVVPKSPLSDVSESDVGTGS 275  
DB 1477 RTSIGSSSSDEYKQK-----DSQSGE-----BEDFIRKQ-IIEAGADEASGS 1519  
QY 276 KNDKNLIRQE---SCLGNSPPFEKES-----EPSPMDVDNSKNSCQDS 316  
DB 1520 E-DEEFIRSQLKEISGVGESQKREBAKGGKGVAGKRRRLTKRSTSFDDAGRRHSWD 1578  
QY 317 EADEETSQGFDEQEDGSSSQATANKPS-----RFOARDADIEFRKRYSTKGGV 364  
DB 1579 E-DDEI---FDESPELKPRETKSQSEELVAVGGGLRFTKIENLSTIADKYSESSQK 1634  
QY 365 R--LHFPQE-----GGESRT---GMNDLNK-----LPGNISSLNVECRNSKH 403  
DB 1635 KTIILYFDEEPELEMESLTDSPDRSRGEGSSILHASSFTPGTSPTSVSLDESDSSPSH 1694  
QY 404 GKDSK-----ITDHLMLPKAEDR-----RKQWETKHORTKI----- 439  
DB 1695 KKGESQQRKARHRSHGPIPLPTIEDSEBEELREBEELLKQEKQELQEQKSSSKS 1754

QY 440 -----PKYVPHLSPOKMLGTPTEEMRMPCGIRLPLRPS-ANHTVT 483  
 Db 1755 KKQDELRAQRERRPKTPSNLPIED--ASTEELRQAEME---ELHRSSSEYSPS 1909  
 QY 484 IRVDLLRAGVVPKPPFHYKDLMD-NKHVMPCSEQLNYPVEDENGERTAGSRWELIQTA 542  
 Db 1810 IESD-----PGFEISPEKIEVQVKYKLPAAVSLYPTDQSVNQKEG-----VQKA 1857  
 QY 543 LNKFTRPQNLKDALIKYNVAYSKWDFAL--IDFWDKVLERAE-----AQHLYQSI 593  
 Db 1858 -----LKSAREMYEEMQPKHYKAPPAANERDEVEFEKPELYGGMLIEDIYESL 1907  
 QY 594 LPD 596  
 Db 1908 VED 1910

RESULT 5  
 ATRX\_PANTR STANDARD; PRT; 2492 AA.  
 AC QYQW4; 15-MAR-2004 (Rel. 43, Created)  
 DT 15-MAR-2004 (Rel. 43, last sequence update)  
 DT 15-MAR-2004 (Rel. 43, last annotation update)  
 DE Transcriptional regulator ATRX (X-linked helicase II) (X-linked  
 DE nuclear protein) (XNP).  
 GN ATRX.  
 OS Pan troglodytes (Chimpanzee).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.  
 OC NCBI\_TaxID=9598;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22763540; PubMed=12777533;  
 RA Kitano T., Schwarz C., Nickel B., Paabo S.;  
 RT "Gene diversity patterns at 10 X-chromosomal loci in humans and  
 RT chimpanzees";  
 RL Mol. Biol. Evol. 20:1281-1289(2003).  
 CC -!- FUNCTION: Could be a global transcriptional regulator. Modifies  
 CC gene expression by affecting chromatin. May be involved in brain  
 CC development and facial morphogenesis.  
 CC -!- SUBUNIT: Probably binds EZH2. Binds annexin V in a calcium and  
 CC phosphatidylcholine/phosphatidylserine-dependent manner (By  
 CC similarity).  
 CC -!- SUBCELLULAR LOCATION: Nuclear. Associated with pericentromeric  
 CC heterochromatin during interphase and mitosis, probably by  
 CC interacting with HPI.  
 CC -!- SIMILARITY: Belongs to the SNF2/RAD54 helicase family.  
 CC -!- SIMILARITY: Contains 1 PHD-type zinc finger.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch))  
 CC -----  
 CC EMBL; AB102642; BAC01111.1; -;  
 DR InterPro; IPR001410; DEAD.  
 DR InterPro; IPR001650; Helicase\_C.  
 DR InterPro; IPR000330; SNF2\_N.  
 DR Pfam; PF00271; Helicase\_C7\_1.  
 DR Pfam; PF00176; SNF2\_N; 1.  
 DR SMART; SM00487; DEXDC; 1.  
 DR SMART; SM00490; HELICC; 1.  
 DR DNA repair; Nuclear protein; DNA-binding; Helicase; ATP-binding;  
 KW Zinc-finger.  
 FT ZN\_FING 220 268 PHD-TYPE  
 FT NP\_BIND 1594 1601 ATP (POTENTIAL).  
 FT SITE 1719 1722 DEGH BOX.  
 FT DOMAIN 745 750 POLY-SER.  
 FT POLY-SER 745 750  
 FT DOMAIN 1151 1156 POLY-SER.

FT DOMAIN 1166 1169 POLY-LYS.  
 FT DOMAIN 1202 1206 POLY-SER.  
 FT DOMAIN 1259 1266 POLY-ASP.  
 FT DOMAIN 1443 1466 POLY-GLU.  
 FT DOMAIN 1499 1502 POLY-GLU.  
 FT DOMAIN 1929 1939 POLY-LYS.  
 FT DOMAIN 1941 1948 POLY-SER.  
 FT DOMAIN 2222 2225 POLY-LYS.  
 FT DOMAIN 2262 2265 POLY-GLU.  
 FT DOMAIN 2420 2425 POLY-GLN.  
 SQ SEQUENCE 2492 AA; 282567 MW; 5697BAC7D37400DC CRC64;  
 Query Match 3.1%; Score 162; DB 1; Length 2492;  
 Best Local Similarity 19.1%; Pred. No. 0.19;  
 Matches 203; Conservative 158; Mismatches 358; Indels 344; Gaps 55;  
 QY 75 KQTIITWMDTKIGKTAESLDSKEN--NNTIESMSSVQKDNFQHNVEKLVNVSQ 132  
 Db 1136 ERENLSKRNTEIKIQSGSSS-DAEESDNNKKQQTSSKKAVIVKCKRNSLRTSTK 1194  
 QY 133 SLDSLSTKSTOYLNQHTAAMCKWQNEGKHTQLLESEPOITVLVPEQ---FSNANIDR 189  
 Db 1195 RKQADIITSSSSDIEDDDQNSI---GEGSSDEQKI--KEVTENLVLSHTGFCQSSGDE 1248  
 QY 190 S-----PQNDPHSDTDSEENRDNOOFTLVK--LANAKQTTEDEHAREAKHQKCSKS 240  
 Db 1249 ALSKSVPTVTDDDDDDNNPENRIAKMLLEEIKANLSSDESDGSSDDEPEEGKRTGKQNE 1308  
 QY 241 CHGECACQOQDEIDVPKSPSDVGSSEDTGSGKND---NKLIRQESCLGNSPP-FEK 296  
 Db 1309 ENFGDEEAKQVNS-----ESDSDEE---SKPRYRHLRLHKLTVSGESGEEK 1356  
 QY 297 ESEPESEPMV---DNSKNSCQDSEADEETSPGDEQEDGSSSQATANKPSFQARDADIEF 353  
 Db 1357 KTFPEKHEKVGKRNRRKVSSEDSDFQESGVSEEV--SESEDEQRPRTSARAKALEE 1414  
 QY 354 RKR-YSTKGGEVRLHPQFEGG-----ESRTGMDDLNAKLP 387  
 Db 1415 NQRSYQKKRRRIKQVEDSSSENKSNSEBEEKEEKEEKEEKEEKEEKEEKEEKEEKEE 1473  
 QY 388 GNISLNVCEKNSKQHGK--KDSKI--TDHLMLPKADEKKEQWETKHQTEKIPKYPV 444  
 Db 1474 G-----KGRKTKIRKLKDKLTETONALKEEERKRIAEERER--EKLREIVIE 1522  
 QY 445 -PHLSPPKWLGTPI-----EEMRRMPRCGIRLPLRPSANHTVTIRVDLLRAGEVKKPF 498  
 Db 1523 IEDASPTKPIITTKVLVDEDEETKE-----PLVQVHRNMVVKL--KP- 1562  
 QY 499 PTHYKD---LWQNKHVMPCEQONLYPVEDENGERTAGSRWELIQTALLNKTRPQNLIK 554  
 Db 1563 --HQVDGVQPMWD-----CCCESV---KTKKSPGSGCILAHCMGLGK---TLQ 1603  
 QY 555 DAILKYNVAYSKKWDF-TALI-----DFWDKVLSEAEQAHLQYQSLPDMVK 599  
 Db 1604 VSFHLTVLLCDKLDSTALVCPNLNWNWNEFEKMQGLKDD----- 1649  
 QY 600 IALCLPNICTQPIPLAKQKMNHSITMSQEQIASLANAFACTFPRRNAMKMSYSSYSDI 659  
 Db 1650 -KLEVSELAIVKEP---QERSYMLQWQEDGVWIIIG-----YEMV--- 1686  
 QY 660 NFNRLFEGRSSRKPEKLTILFCYFRRRYTEKKPTGLVTFTRQSLDFEWECEKPL----- 715  
 Db 1687 --RNLAQGRNV-KSRKLKEIF--NKALVDPGDFVVCDEGHILKN--EASAVSKAMNSIR 1739  
 QY 716 TRLHVTVYEGT-----TEENGQGMLOVDFAFRV-----GGGVTS----- 749  
 Db 1740 SRRRIITGTPQLNNLIEYCMVNFINKENLLGSIK-EFRNRFNFIQNGQCASTWVDVR 1798  
 QY 750 -----AGLVQEE-----IRFL-----INPELIISRLFTFVLDHNECLI 782  
 Db 1799 VMKRAHILYEMLAGCVQRKYDTALTKEFLPPKHEYVLAVRMTSIOCKLYQVYLDH----- 1853  
 QY 783 ITGTEQYSE-----YTGYAETYRW-----SRSHEDGSEDDCCERRCTE 820

entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

[illegible]

Db 1559 ---CCCESV-----EKKTKSPGSCILAHCMGLGK-----TLQVVSFLHTVLICDKLDFS 1605  
QY 571 TALIDF-----WDKVLBAEAQHLYQSILPDMVKIALCLPNCITQPIPLKQKXHSI 623  
Db 1606 TALVVCPLNTALNWMNEFEK-----WQGLNDNEKLE--VSELATVGRP---QERSYML 1654  
QY 624 TMSQEQIASILANAFCTCFPRNAKMKSEYSSYPDINFNRLPFGSRSRPEKLTILFCYF 683  
Db 1655 QRWQEDGGVMIIG-----YEMY-----RNLAGRVV-KSKLKDIF--N 1690  
QY 684 RRVTEKPTGLVTFTRQSLDFPEWERCPL-----TRLHVTYEGT----- 725  
Db 1691 KALVDPGDFVDCDEGHILKN--EASAVSKAMNSIKSRRIILGTGPLQNNLIEYHCWN 1748  
QY 726 IEENGOMLOVDPAHFV-----GGGVTS-----AGLVQEB----- 756  
Db 1749 FIKENLIGSIK-EPRNRFINPIQNGQCADTWDVVRVKKRAHLYEMLAGCVQRKDYTA 1807  
QY 757 -IRFL-----INPELLISRLFTVLDHNECLITG---TE-----QYSE 791  
Db 1808 LTRFLPPKHEYVLAVRMATLQCKLYQYLDH-----LTGVNSTEGRGKAGAKLFDQFQ 1862  
QY 792 YTGAYTYRM-----SRSHEDSGSDCCERRCTEIVADLHFRYRLQDFVPERMRE 844  
Db 1863 MLSRIWHPCLQLDYISKENKGVFDEDSMD-----EFIASDSDETSKSLSS--DEKKPK 1916  
QY 845 LNKAYCGFLRPGVSSENLAVATGWCAGGAFGDPARLKALIQILAAAAERDVVYFTFGD 904  
Db 1917 GKKGKDDSSSGSGSDN--DVEVIKWNRSRGG-----GD 1950  
QY 905 SEL--MEDIYSMHIFLTKVTDGVYKLLRLYYNEECNCTGPGDIKLYPFYTHAVES 962  
Db 1951 GNMDDTGNFVSUKLDESKTT-----STSNPSSPAPDW-----YKDFYDT 1992  
QY 963 CAETADHSGQ 972  
Db 1993 DAEVLEHSGK 2002  
RESULT 7  
ATRX HUMAN STANDARD; PRT: 2492 AA.  
ID ATRX HUMAN STANDARD; PRT: 2492 AA.  
AC P46100; P51068; Q15886; Q9H021; Q9NTS3;  
DT 01-NOV-1995 (Rel. 32, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 15-MAR-2004 (Rel. 43, Last annotation update)  
DE Transcriptional regulator ATRX (X-linked helicase II) (X-linked  
DE nuclear protein) (XNP) (Znf-HX).  
GN ATRX OR RAD54L OR XH2.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A. (ISOFORMS 1; 2; 3; 4 AND 5), VARIANTS SER-596;  
RP GLU-740 AND SER-1860, AND VARIANTS ATR-X.  
RX MEDLINE=97123494; PubMed=8968741;  
RA Picketts D.J., Higgs D.R., Bachoo S., Blake D.J., Quarrell O.W.J.,  
RA Gibbons R.J.;  
RT "ATRX encodes a novel member of the SNF2 family of proteins: mutations  
RT point to a common mechanism underlying the ATR-X syndrome.";  
RL Hum. Mol. Genet. 5:1899-1907(1996).  
RN [2]  
RP SEQUENCE FROM N.A. (ISOFORMS 2 AND 4).  
RX MEDLINE=97386582; PubMed=9244431;  
RA Villard L., Lossi A.-M., Cardoso C., Proud V., Chiaroni P.,  
RA Collea L., Schwartz C., Fontes M.;  
RT "Determination of the genomic structure of the XNP/ATRX gene encoding  
RT a potential zinc finger helicase";  
RL Genomics 43:149-155(1997).  
RN [3]  
RP SEQUENCE FROM N.A., AND VARIANTS SER-596 AND GLU-740.  
RX MEDLINE=22763540; PubMed=12777533;

RA Kitano T., Schwarz C., Nickel B., Paabo S.;  
RT "Gene diversity patterns at 10 X-chromosomal loci in humans and  
RL chimpanzees";  
RN Mol. Biol. Evol. 20:1281-1283(2003).  
RP [4]  
RP SEQUENCE OF 860-2492 FROM N.A.  
RX MEDLINE=95179111; PubMed=7874112;  
RA Stayton C.L., Dabovic B., Gulisano M., Gez J., Broccoli V.,  
RA Giovanazzi S., Bosso Lasco M., Monaco L., Rastan S., Boncinelli E.,  
RA Bianchi M.E., Consalez G.G.;  
RT "Cloning and characterization of a new human Xq13 gene, encoding a  
RT putative helicase";  
RL Hum. Mol. Genet. 3:1957-1964(1994).  
RN [5]  
RP PRELIMINARY PARTIAL SEQUENCE FROM N.A.  
RX MEDLINE=94214473; PubMed=8162050;  
RA Gez J., Pollard H., Consalez G., Villard L., Stayton C.L.,  
RA Millasseau P., Khrestchatsky M., Fontes M.;  
RT "Cloning and expression of the murine homologue of a putative human  
RT X-linked nuclear protein gene closely linked to PGK1 in Xq13.3";  
RL Hum. Mol. Genet. 3:39-44(1994).  
RN [6]  
RP SEQUENCE OF 2401-2492 FROM N.A., AND VARIANTS ATR-X.  
RX MEDLINE=95211835; PubMed=7697714;  
RA Gibbons R.J., Picketts D.J., Villard L., Higgs D.R.;  
RT "Mutations in a putative global transcriptional regulator cause X-  
RT linked mental retardation with alpha-thalassemia (ATR-X syndrome).";  
RL Cell 80:837-845(1995).  
RN [7]  
RP SEQUENCE OF 1375-2492 FROM N.A.  
RA Pearce A., Chapman J.;  
RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.  
RN [8]  
RP E2H2 BINDING.  
RX MEDLINE=98167853; PubMed=9499421;  
RA Cardoso C., Timmit S., Villard L., Khrestchatsky M., Fontes M.,  
RA Collea L.;  
RT "Specific interaction between the XNP/ATR-X gene product and the SET  
RT domain of the human E2H2 protein";  
RL Hum. Mol. Genet. 7:679-684(1998).  
RN [9]  
RP SUBCELLULAR LOCATION, AND ASSOCIATION WITH PERICENTROMERIC  
RP HETEROCHROMATIN.  
RX MEDLINE=20040663; PubMed=10570185;  
RA McDowell T.L., Gibbons R.J., Sutherland H., O'Rourke D.M.,  
RA Bickmore W.A., Pombo A., Turley H., Gatter K., Picketts D.J.,  
RA Buckle V.J., Chapman L., Rhodes D., Higgs D.R.;  
RT "Localization of a putative transcriptional regulator (ATRX) at  
RT pericentromeric heterochromatin and the short arms of acrocentric  
RT chromosomes";  
RL Proc. Natl. Acad. Sci. U.S.A. 96:13983-13988(1999).  
RN [10]  
RP DISEASE.  
RX MEDLINE=20213147; PubMed=10751095;  
RA Villard L., Fontes M., Ades L.C., Gez J.;  
RT "Identification of a mutation in the XNP/ATR-X gene in a family  
RT reported as Smith-Fineman-Myers syndrome.";  
RL Am. J. Med. Genet. 91:83-85(2000).  
RN [11]  
RP DISEASE.  
RX MEDLINE=22804448; PubMed=12858175;  
RA Gibbons R.J., Pellagatti A., Garrick D., Wood W.G., Malik N.,  
RA Ayyub H., Langford C., Boulwood J., Wainscoat J.S., Higgs D.R.;  
RT "Identification of acquired somatic mutations in the gene encoding  
RT chromatin-remodeling factor ATRX in the alpha-thalassemia  
RT myelodysplasia syndrome (ATMDS).";  
RL Nat. Genet. 34:446-449(2003).  
RN [12]  
RP VARIANT ATR-X SER-1713.  
RX MEDLINE=97196774; PubMed=9043863;  
RA Villard L., Lacombe D., Fontes M.;  
RT "A point mutation in the XNP gene, associated with an ATR-X phenotype  
RT without alpha-thalassemia.";

RL Eur. J. Hum. Genet. 4:316-320(1996).  
 RN [13]  
 RP VARIANT JM GLN-2131.  
 RX MEDLINE=96224392; PubMed=8630485;  
 RA Villard L., Gecz J., Mattei J.-F., Fontes M., Saugier-Verber P.,  
 RA Munnich A., Lyonnet S.,  
 RT "XNP mutation in a large family with Juberg-Marsidi syndrome.";  
 RL Nat. Genet. 12:359-360(1996).  
 RN [14]  
 RP VARIANTS ATR-X.  
 RX MEDLINE=97467722; PubMed=9326931;  
 RA Gibbons R.J., Bachoo S., Picketts D.J., Aftimos S., Aasenbauer B.,  
 RA Bergoffen J., Berry S.A., Dahl N., Fryer A., Keppler K., Kurosawa K.,  
 RA Levin M.L., Masuno M., Neri G., Pierpont M.E., Slaney S.F.,  
 RA Higgs D.R.;  
 RT "Mutations in transcriptional regulator ATRX establish the functional  
 RT significance of a PHD-like domain.";  
 RL Nat. Genet. 17:146-148(1997).  
 RN [15]  
 RP VARIANT ATR-X LEU-246.  
 RX MEDLINE=20123062; PubMed=10660327;  
 RA Fichera M., Romano C., Castiglia L., Failla P., Ruberto C., Amata S.,  
 RA Greco D., Cardoso C., Fontes M., Ragusa A.;  
 RT "New mutations in XNP/ATR-X gene: a further contribution to  
 RT genotype/phenotype relationship in ATR/X syndrome.";  
 RL Hum. Mutat. 12:214-214(1998).  
 RN [16]  
 RP VARIANT SHS LYS-1742.  
 RX MEDLINE=99347960; PubMed=10417298;  
 RA Lassi A.-M., Millan J.M., Villard L., Orellana C., Cardoso C.,  
 RA Prieto F., Fontes M., Martinez P.;  
 RT "Mutation of the XNP/ATR-X gene in a family with severe mental  
 RT retardation, spastic paraplegia and skewed pattern of X inactivation:  
 RT demonstration that the mutation is involved in the inactivation  
 RT bias.";  
 RL Am. J. Hum. Genet. 65:558-562(1999).  
 RN [17]  
 RP VARIANT CWS THR-2050.  
 RX MEDLINE=99326061; PubMed=10398237;  
 RA Abidi F., Schwartz C.E., Carpenter N.J., Villard L., Fontes M.,  
 RA Curtis M.;  
 RT "Carpenter-Waziri syndrome results from a mutation in XNP.";  
 RL Am. J. Med. Genet. 85:249-251(1999).  
 RN [18]  
 RP VARIANTS ATR-X GLU-175; 178-VAL--LYS-198 DEL; SER-190; PRO-219;  
 RX LEU-246 AND CYS-249.  
 RX MEDLINE=99219535; PubMed=10204841;  
 RA Villard L., Bonino M.-C., Abidi F., Ragusa A., Belougne J.,  
 RA Lassi A.-M., Seaver L., Bonnefont J.-P., Romano C., Fichera M.,  
 RA Lacombe D., Hanauer A., Philip N., Schwartz C.E., Fontes M.;  
 RT "Evaluation of a mutation screening strategy for sporadic cases of  
 RT ATR-X syndrome.";  
 RL J. Med. Genet. 36:183-186(1999).  
 RN [19]  
 RP VARIANTS ATR-X SER-179; LEU-190; ILE-194; CYS-246; PHE-1553; SER-1645  
 AND CYS-1847.  
 RX MEDLINE=20451413; PubMed=10995512;  
 RA Wada T., Kubota T., Fukushima Y., Saitoh S.;  
 RT "Molecular genetic study of Japanese patients with X-linked alpha-  
 RT thalassemia/mental retardation syndrome (ATR-X).";  
 RL Am. J. Med. Genet. 94:242-248(2000).  
 RN [20]  
 RP VARIANT ATR-X MET-1621.  
 RX MEDLINE=22108790; PubMed=12116232;  
 RA Yntema H.G., Poppelars F.A., Derksen E., Oudakker A.R.,  
 RA van Roosmalen T., Jacobs A., Obbema H., Brunner H.G., Hamel B.C.J.,  
 RA van Bokhoven H.;  
 RT "Expanding phenotype of XNP mutations: mild to moderate mental  
 RT retardation.";  
 RL Am. J. Med. Genet. 110:243-247(2002).  
 CC -1- FUNCTION: Could be a global transcriptional regulator. Modifies  
 CC gene expression by affecting chromatin. May be involved in brain  
 CC development and facial morphogenesis.

CC -1- SUBUNIT: probably binds EZH2. Binds annexin V in a calcium and  
 CC phosphatidylcholine/phosphatidyleserine-dependent manner (by  
 CC similarity).  
 CC -1- SUBCELLULAR LOCATION: Nuclear. Associated with pericentromeric  
 CC heterochromatin during interphase and mitosis, probably by  
 CC interacting with HPI.  
 CC -1- ALTERNATIVE PRODUCTS:  
 CC Event-Alternative splicing; Named isoforms=5;  
 CC Name=4;  
 CC IsoId=P46100-1; Sequence=Displayed;  
 CC Name=1;  
 CC IsoId=P46100-2; Sequence=VSP\_000575;  
 CC Name=2;  
 CC IsoId=P46100-3; Sequence=VSP\_000574;  
 CC Name=3;  
 CC IsoId=P46100-4; Sequence=VSP\_000576;  
 CC Name=5;  
 CC IsoId=P46100-5; Sequence=VSP\_000574, VSP\_000576;  
 CC -1- TISSUE SPECIFICITY: Ubiquitous.  
 CC -1- DISEASE: Defects in ATRX are the cause of X-linked alpha-  
 CC thalassemia/mental retardation syndrome (ATR-X) [MIM:301040]. ATR-  
 CC X is an X-linked disorder comprising severe psychomotor  
 CC Query Match 3.1%; Score 161; DB 1; Length 2492;  
 CC Best Local Similarity 19.1%; Pred. No. 0.22;  
 CC Matches 203; Conservative 158; Mismatches 358; Indels 344; Gaps 55;  
 QY 75 KOKTTSMDTKGKTAESLSDSKEN--NNTRIEMSSVQKDNFYQHNVKLVNVSQ 132  
 DB 1136 ERNLSSKNTKEIQSGSSS--DAEESDNNKKKQRTSKKAVTVKKKNSLRTSTK 1194  
 QY 133 SLDKSLTEKSTQYLQHQTAAKCKQNEKHTEQLLSEPTQTVTLVPEQ---FSNANIDR 189  
 DB 1195 RQADITSSSSSDIEDDDQNSI---GGSSDEQKI--KPVTEIIVLSSHTGFCQSSGDE 1248  
 QY 190 S-----PQNDHSDTSEENRDNQOFLTTVK--LANAKOTTEDEHABAKSHQKCSKS 240  
 DB 1249 ALSKSVPTVDDDDNDPENRIAKMLLEIKANLSDSDGSSDDEPEEGKRTQKQNE 1308  
 QY 241 CHGEDCASCQDEIDVVPKPLSVGSESDVGTGSKND--NKLIRQESCLGNSPP-FEK 296  
 DB 1309 ENPGDEEAKQVNS-----ESDSDSE--SKPRYRHLRLKHLTVSDGESGEK 1356  
 QY 297 ESEPSPMDV--DNSKNSCQSEADEETSPGDEQEDGSSQTSANKPFRQARDADIEF 353  
 DB 1357 KTKPKHEKVGKRRKVSSESDSDFOESGVSEV--SESEDEQRPRTRSAKAELEE 1414  
 QY 354 RKR-YSTKGGEVRLHPQFEGG-----ESRTGMNDLNAKLP 387  
 DB 1415 NQRSYKQKKRRRIKQVEDSSSENKSNSEEEEEEEEEEEEEEEDEED--DSKSP 1473  
 QY 388 GNISSLNVECRNSKQHGK--KDSKI--TDHLMRLPKAEDRRKQWETKHQTERKIPKYP 444  
 DB 1474 G-----KGRKTRKILKODKLTQNAKSEEEERKRAERER--EKLEEVIE 1522  
 QY 445 -PHLSPPKWLGTPI-----EEMRMPCGIRLPLRPSANHTVTTRVLLRAGEVKKPF 498  
 DB 1523 IEDASPTKCPITKLVLDEDEETKE-----PLQVHRNMVVKL-----KP- 1562  
 QY 499 PTHYKD----LMDNKHVMPKSEQNLYPVEDENGERTAGRWELIQTALLNKTRQNLIK 554  
 DB 1563 --HOVDGVQFMD-----CCCESV-----KTKKSPGSGCILAHCMGLGK-----TLQ 1603  
 QY 555 DALKKNVAYSKWDF-TALI-----DFWDKVLFEAEAAQHLVQSILPDMVK 599  
 DB 1604 VWSFLHTVLCDKLDFTALVVCPLNTALNMWNEFEKQWGLKDD------ 1649  
 QY 600 IALCLPNICTQPIPLLKQKNHSITWSQEOIASILANAFCTPPRRNAKMKSYSSYPDI 659  
 DB 1650 -KLEVEELATVKP---QERSYMLQWQEDGGVMIIG-----YEMV--- 1686  
 QY 660 NFNRLEFGRSSRKPEKLTLCFYFRVRTEKPKTGLVTFTRQSLDFPWERCEKPL----- 715

```

Db 1687 --RNLQGRNV-KSRKLKEIF--NKALVDPGDFVVCDEGHILKN--EASAVSKAMNSIR 1739
QY 716 TRLHVITYEFT-----TEENQGMQLQVDFANRFV-----GGGVTS----- 749
Db 1740 SRRRIILGTPLQNLLEIYHGMWNFKENILGSIK-EFRNRFINPIQNGQCADSTWVDVR 1798
QY 750 -----AGLVOEE-----IRFL-----INPELLISRLFEVLHDHNECLI 782
Db 1799 VMKRAHILYEMLAGCVQRDYALTATKFLPPKHEVYLAVMWTSIQCKLYQYLDH----- 1853
QY 783 ITGTEQYSE-----YTGYARTYRW-----SRSHEDGSRDDCERRCTE 820
Db 1854 LTGVGNNSGGRGKAGAKLQDFQMLSLIWIHPWCLQLDVISEKNGKGYFDEDSMD--E 1909
QY 821 IVALDLHFRRYL--DOFVPEKMRRELNKAYGGLRPGVSENLSAVATGN----- 869
Db 1910 FIASDSDETSMSLSDDYTKKKKGKKGK-----KDSGSGSGSDNDVEVIK 1957
QY 870 WGGCAGFGDARLKALIQILAAAAARDVWYFTFGDSELMRDIVSMHIFLTERKLTVGDVY 929
Db 1958 WNSRSGGGE-----GNVDETGNPSVSKLESKAT----- 1989
QY 930 KLLRLYYNEECRNGSTPGPIKLYFFIYHAVESCAETADHSQ 972
Db 1990 -----SSSNFSSPAPDW-----YKDFVTDADAELVLEHSGK 2019

RESULT 8
ATRX PONPY
ID ATRX PONPY STANDARD; PRT; 2492 AA.
AC QYQWQ3;
DT 15-MAR-2004 (Rel. 43, Created)
DT 15-MAR-2004 (Rel. 43, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Transcriptional regulator ATRX (X-linked helicase II) (X-linked
DE nuclear protein) (XNP).
GN ATRX.
OS Pongo pygmaeus (Orangutan).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pongo.
OC NCBI_TaxID=9600;
RN [1]
RP MEDLINE=22763540; PubMed=12777533;
RA Kitano T., Schwarz C., Nickel B., Paabo S.;
RT "Gene diversity patterns at 10 X-chromosomal loci in humans and
RT chimpanzees.";
RL Mol. Biol. Evol. 20:1281-1289(2003).
CC -!- FUNCTION: Could be a global transcriptional regulator. Modifies
CC gene expression by affecting chromatin. May be involved in brain
CC development and facial morphogenesis.
CC -!- SUBUNIT: Probably binds EZH2. Binds annexin V in a calcium and
CC phosphatidylcholine/phosphatidylserine-dependent manner (By
CC similarity).
CC -!- SUBCELLULAR LOCATION: Nuclear. Associated with pericentromeric
CC heterochromatin during interphase and mitosis, probably by
CC interacting with HPI.
CC -!- SIMILARITY: Belongs to the SNF2/RAD54 helicase family.
CC -!- SIMILARITY: Contains 1 PHD-type zinc finger.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AB102643; BAC81112.1; -.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR001650; Helicase_C.
DR InterPro; IPR000330; SNF2_N.
DR Pfam; PF00271; helicase_C; 1.

```

```

DR Pfam; PF00176; SNF2_N; 1.
DR SMART; SM00487; DEXDC; 1.
DR SMART; SM00490; HELIC; 1.
KW DNA repair; Nuclear protein; DNA-binding; Helicase; ATP-binding;
KW Zinc-finger.
FT ZN FING 220 268 PHD-TYPE.
FT NF BIND 1594 1601 ATP (POTENTIAL).
FT SITE 1719 1722 DEGH BOX.
FT DOMAIN 745 750 POLY-SER.
FT DOMAIN 1151 1156 POLY-SER.
FT DOMAIN 1166 1169 POLY-LYS.
FT DOMAIN 1202 1206 POLY-SER.
FT DOMAIN 1259 1266 POLY-ASP.
FT DOMAIN 1443 1466 POLY-GLU.
FT DOMAIN 1499 1502 POLY-GLU.
FT DOMAIN 1929 1939 POLY-LYS.
FT DOMAIN 1941 1948 POLY-SER.
FT DOMAIN 2222 2225 POLY-LYS.
FT DOMAIN 2262 2265 POLY-GLU.
FT DOMAIN 2420 2425 POLY-GLN.
SQ SEQUENCE 2492 AA; 282614 MW; 48EC97106D0CEF20 CRC64;

Query Match 3.1%; Score 161; DB 1; Length 2492;
Best Local Similarity 19.1%; Pred. No. 0.22;
Matches 203; Conservative 158; Mismatches 358; Indels 344; Gaps 55;

QY 75 KQVITISMDTKIGKTAESLSLDSKEN--NNTRIESMMSSVQKDNFYQHNVEKLVNVSQ 132
DB 1136 ERRNLSSKRNTEIQSGSSS--DAEESSEDNKKKQRTSSKKAVIVKEKKRSLRTSTK 1194
QY 133 SLDSKSLTEKSTQYLNQHOTAAKCKQWQNEGHKHTQLLESEPTVTLVPEQ---FSNANIDR 189
DB 1195 RKQADITSSSSSDIEDDDQNSI---GEGSSDQKI--KPVTEINLVLSHTTGFQCSGDE 1248
QY 190 S-----PONDHSDDTDSENRDNQOFLTIVK--LANAKQTTDEDEHAREAKSHQKCSKS 240
DB 1249 ALSKSVPTVTDNDDDDNDPENRIAKMLLEEIKANLSSDESDGSDDEPFGKKTGTGQNE 1308
QY 241 CHGEDCASQODELDVVPKSPLDVSGSDVGTGSKND---NKLIRIESCLGNSPP--FEK 296
DB 1309 ENPGDEEAKQVNS-----ESDSDEE---SKPRYRHLRLHKLIVSDGESGEK 1356
QY 297 ESEPESEPMQV---DNSKNSQDSEADEETSFGDFQEDSGSSSOTANKPSRFOQADIEF 353
DB 1357 KTRPKRHEKVGKRRNRKVSSEDSSEDSDFQSGVSEEV--SESEDEQRPRTSRAKAELEE 1414
QY 354 RKR-YSTKGGEVRLHFQEGG-----ESRTGMNDLNAKL 387
DB 1415 NORSYKOKKRRRIKVOEDSSSENKNSSEEEKEKEKEKEKEKEKEKEKEKEKEKEKEKE 1473
QY 388 GNISLNVCEKNSKQHGK--KDSKI-TDHLMLRPLKABDRKQEWETKHQTERKIPKYP 444
DB 1474 G-----KGRKIRKILKDDKLRTEQNALKEEERKRKRAERER--EKLEBVE 1522
QY 445 -PHLSPDKMWLGTPI-----EEMRMPCRGIRPLLRPSANHVTVIRVLLRAGEVDPKP 498
DB 1523 IEDASPTKCPITTKVLDEDEETKE-----PLVQVHRNMVTKL-----KP- 1562
QY 499 PTHYKD-----LWQNKHVMPQSEONLYPVEDENGERTAGSRWELIQTALLNKFTRPNL 554
DB 1563 --HQVDGVQVMDW-----CCCESV-----KTKKSPGSGCILLAHCMGLCK--TLQ 1603
QY 555 DAILKNVAVSKKQWDF--TALI-----DFWDKVLBEAAEQHLYQSILPDMVK 599
DB 1604 VVSFLHVLLCDKLDLFTALVCPPLNTALNMWNEFEKQWQSLKDD----- 1649
QY 600 IALCLPNICTQPIPLLKQKNHSTMSQEQIASLANAFAFFCTPPRRNAKMSSEYSSPDI 659
DB 1650 -KLEVSELATVTRP---QERSYMLQRWQEDGVNMIIG-----YEMV--- 1686
QY 660 NFNRLFEGRSRRKPEKLTFLFCYFRRVTEKKPTGLVTFTRQSLDFPWERCEKPL--- 715
DB 1687 --RNLAQGRNV-KSRKLKEIF--NKALVDPGDFVVCDEGHILKN--EASAVSKAMNSIR 1739

```



```

QY 716 TRLVHYEGT-----TEENGQMLQVDPAARFV-----GGVTS-----749
DB 1740 SRRRIILGTPLQNNLIEHYCMWNFKENLLGSIK-EFRNRFINPQNGQCADSTWVDR 1798
QY 750 -----AGLVOEE-----IRFL-----INPELLIIRLFTVLDHNECLI 782
DB 1799 VMKRAHIIYELAGCVQRKYDTALTFFPKPKHEYVLAVMTSIOCKLQYYLDH-----1853
QY 783 ITGTEQYSE-----YTGAYETVW-----SRSHEDGSEDDCHERCVE 820
DB 1854 LTGVGNNSGGRGKAGAKLFQDPQMLSRITWPCQLQLDYISKENGYFDEDSMD-----E 1909
QY 821 IVALDALHFERYL--DQFVPEKWRRELNKAYCGFLRPGVSSENLSAVATGN-----869
DB 1910 FIASDSTSMSSDDYTKKKKKKKKK-----KSSSGSGSDNDVEIKV 1957
QY 870 WCGAGFGDARLKALIQILAAAAAERDVVYFTFGDSLMRDYISMIFITERKLTVGDVY 929
DB 1958 WNSRSGGGE-----GNVDETGNPNPSVLKLEESKAT-----1989
QY 930 KLLRLYNEECNCSPTGPDPIKLYPIYHVESCAETADHSQ 972
DB 1990 -----SSNPSPPAPDW-----YKDFVTDADAELVLEHSGK 2019

RESULT 9
BMS1_YEAST
ID BMS1_YEAST STANDARD; PRT; 1183 AA.
AC Q08965;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Ribosome biogenesis protein BMS1.
GN BMS1 OR YPL217C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / AB972;
RA Bussey H., Storms R.K., Ahmed A., Albermann K., Allen E., Ansoerge W.,
RA Araujo R., Aparicio A., Barrell B.G., Badcock K., Benes V.,
RA Botstein D., Bowman S., Bruckner M., Carpenter J., Cherry J.M.,
RA Chung E., Churcher C.M., Coster F., Davis K., Davis R.W.,
RA Dietrich F.S., Delius H., Dipaolo T., Dubois E., Duesterhoeft A.,
RA Duncan M., Floeth M., Fortin M., Friesen J.D., Fritz C., Goffeau A.,
RA Hall J., Hebling U., Heumann K., Hilbert H., Hillier L.W.,
RA Hunicke-Smith S., Hyman R., Johnston M., Kalman S., Kleine K.,
RA Komp C., Kurdi O., Lashkari D., Lew H., Lin A., Lin D., Louis E.J.,
RA Marathe R., Messenguy F., Mewes H.-W., Mirtipati S., Moestl D.,
RA Mueller-Auer S., Namath A., Newtich U., Oefner P., Pearson D.,
RA Petel F.X., Pohl T.M., Purnelle D., Schafer M., Scharfe M.,
RA Scherens B., Schramm S., Schroeder M., Sidcu A.M., Tettelin H.,
RA Urretarazu L.A., Ushinsky S., Vierendeels F., Vissers S., Voss H.,
RA Walsh S.W., Wambutt R., Wang Y., Wedler E., Wedler H., Wilmott E.,
RA Zhong W.W., Zollner A., Vo D.H., Hani J.;
RT "The nucleotide sequence of Saccharomyces cerevisiae chromosome XVI.";
RL Nature 387:103-105(1997).
RN [2]
RP FUNCTION.
RC MEDLINE=21449426; PubMed=11565748;
RA Wegierski T., Billy E., Nasr F., Filipowicz W.;
RT "Bmslp, a G-domain-containing protein, associates with Rclp and is
RT required for 18S rRNA biogenesis in yeast.";
RL RNA 7:1254-1267(2001).
CC -!- FUNCTION: May act as a molecular switch during maturation of the
CC 40S ribosomal subunit in the nucleolus. The depletion of BMS1
CC interferes with processing of the 35S pre-rRNA at sites A0, A1,
CC and A2, and the formation of 40S subunits.
CC -!- SUBUNIT: Associates with RCL1.

```

```

CC -!- SUBCELLULAR LOCATION: Nuclear; nucleolar.
CC -!- SIMILARITY: Belongs to the BMS1 family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; Z73573; CAA97932.1; -
CC PIR; S65236; S65236.
CC SGD; S0006138; BMS1.
CC GO; GO:0005737; C:cytoplasm; IDA.
CC GO; GO:0005730; C:nucleolus; IDA.
CC GO; GO:0006365; P:35S primary transcript processing; IDA.
CC GO; GO:0042255; P:ribosome assembly; IDA.
CC InterPro; IPR003593; AAA ATPase.
CC InterPro; IPR007034; DUF663.
CC Pfam; PF04950; DUF663; 1.
CC SMART; SM00382; AAA; 1.
KW Ribosome biogenesis; Nuclear protein; ATP-binding.
FT NP_BIND 76 83 ATP (POTENTIAL).
SQ SEQUENCE 1183 AA; 135570 MW; 9A337F1EE080F21D CRC64;

Query Match 3.1%; Score 159.5; DB 1; Length 1183;
Best Local Similarity 18.3%; Pred. No. 0.1; Indels 307; Gaps 36;
Matches 156; Conservative 116; Mismatches 274;

QY 33 SRQRRVLDPKD-----AHQFRVPSPSPACVPGQ-AGQHRGSAISL 72
DB 341 TRRRKELDDKDLIYAPMSDVGVLMDKDAVIDIGKNKEEPSFVPGQGRGEKLMTGL 400
QY 73 VFQKTIITSMDTKGI-----KTAESLSLSKENNTRIE-----SMWSSVQKD 116
DB 401 QSVESQIAEKFDGVLQGLFNGTSLHTEHVAHEGMD-VESGESIEDDEGSKGRTSLRKP 459
QY 117 NFYQHNV-EKLVNVSLSLDKSLTEKSTOYLNOHQHTAAMCKWQNECKHTEQLLESPTV 175
DB 460 RIYKGVQVEDADIDNLPSEDE-----PYTND-----DDVQDSFPRMV 497
QY 176 TLVPEQFSNANIDRSPQNDHDSPTSEENRDNQQLTITVKLANAKQTTEDEHA-----228
DB 498 EI--DFNNTGSGAELKLALETDSFESEDEFSWERTA--ANKLKKTESKKTWNIGKL 552
QY 229 --REAKSHOKSKSGHPGDDCHASCQODEIDVVPKSPSLSDVGSDEV-----GTGSKN 277
DB 553 IYMDNISPEECIRWR-GEEDDS--KDESDI-----BEDVDDDFRKKDGTITKE 599
QY 278 DNK--LIROESCLGNSPPPEKESEPESPMDV-----DNSKNSQDSDEADEETS 323
DB 600 GAKDHAVDLEKFPVPTDFTEKLAKWKWSVDAIKERFLGAGILGNDNKTSDSNEGGEELY 659
QY 324 PGFDEQEDGSSQTANKPSRFOARDAD-----IIEPKRY 357
DB 660 GDFELEDGDNPSQEAEDNSDESEDEDEDNEDTGDNSFTNFDABEKKDLTWEQEREMN 719
QY 358 STKGGEVRLHFQEGESGRTGNDLN-----AKLPGNISLIANVEC-----RNS 400
DB 720 AAKKEKLRAQFIEGKFNKEDDENNEYDTWYELQKAKISKQLEINNIYQEWMTPEQRQR 779
QY 401 KQHGKKDSKITDHLMLRP-----KAEDRRKEOME-----429
DB 780 IEGFKAGSVIRIVFEKVPMEFVNFPNPKPIVMGGLLPTEIKFIVKARLR-HRWHKKI 838
QY 430 -----TKHQPTERKIPKVPPLSPDKKMLG---TPIEEMR 463
DB 839 LKTNPLVLSLGWRRFOTLPITVTTDSRTRTRMLKYTPHEITYCNAAFYGLPCSP-----N 893
QY 464 MPRCGIRLPLLPSSAN-----HTVTIRVDLLRAGEVPKPPRTHYKDLWDNKHVK- 512
DB 894 TPFQGVQIVANSDTGNGFFIAATGIVVEEDVNIIEIVKLUKL-VGFP--YKIFKNTAFIKD 950

```

[illegible]

GN SPTAL OR SPTA.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OC NCBI\_TaxID=9606;  
RN [1]\_TaxID=9606;  
RP SEQUENCE FROM N.A.  
RX MEDLINE=90170949; PubMed=1693726;  
RA Sahr K.E., Tobé T., Scarpa A., Laughtinghouse K., Marchesi S.L.,  
RA Linnenbach A.J., Winkelmann J.C., Speicher D.W., Marchesi V.T.,  
RA Curtis P.J., Forget B.G.;  
RT "The complete cDNA and polypeptide sequences of human erythroid  
RT alpha-spectrin.";  
RL J. Biol. Chem. 265:4434-4443 (1990).  
RN [2]  
RP SEQUENCE OF 7-533 FROM N.A., AND VARIANTS EL2 PRO-260; PRO-261 AND  
RP PRO-471.  
RX MEDLINE=90009318; PubMed=2794061;  
RA Sahr K.E., Tobé T., Scarpa A., Laughtinghouse K., Marchesi S.L.,  
RA Agre P., Linnenbach A.J., Marchesi V.T., Forget B.G.;  
RT "Sequence and exon-intron organization of the DNA encoding the alpha  
RT I domain of human spectrin. Application to the study of mutations  
RT causing hereditary elliptocytosis.";  
RL J. Clin. Invest. 84:1243-1252 (1989).  
RN [3]  
RP SEQUENCE OF 7-601.  
RX MEDLINE=84087888; PubMed=6654896;  
RA Speicher D.W., Davis G., Marchesi V.T.;  
RT "Structure of human erythrocyte spectrin. II. The sequence of the  
RT alpha-I domain.";  
RL J. Biol. Chem. 258:14938-14947 (1983).  
RN [4]  
RP SEQUENCE OF 7-125.  
RX MEDLINE=84087887; PubMed=6654895;  
RA Speicher D.W., Davis G., Yurchenco P.D., Marchesi V.T.;  
RT "Structure of human erythrocyte spectrin. I. Isolation of the alpha-  
RT domain and its cyanogen bromide peptides.";  
RL J. Biol. Chem. 258:14931-14937 (1983).  
RN [5]  
RP SEQUENCE OF 320-450 FROM N.A.  
RX MEDLINE=86205962; PubMed=3458204;  
RA Linnenbach A.J., Speicher D.W., Marchesi V.T., Forget B.G.;  
RT "Cloning of a portion of the chromosomal gene for human erythrocyte  
RT alpha-spectrin by using a synthetic gene fragment.";  
RL Proc. Natl. Acad. Sci. U.S.A. 83:2397-2401 (1986).  
RN [6]  
RP PARTIAL SEQUENCE.  
RX MEDLINE=84295638; PubMed=6472478;  
RA Speicher D.W., Marchesi V.T.;  
RT "Erythrocyte spectrin is comprised of many homologous triple helical  
RT segments.";  
RL Nature 311:177-180 (1984).  
RN [7]  
RP IDENTIFICATION OF PROBABLE FRAMESHIFT IN 2407-2418.  
RA Gibson T.J.;  
RN Unpublished observations (MAR-1995).  
RN [8]  
RP REVIEW ON VARIANTS.  
RX MEDLINE=97001215; PubMed=8844207;  
RA Mailliet P., Aloisio N., Morie L., Delaunay J.;  
RT "Spectrin mutations in hereditary elliptocytosis and hereditary  
RT spherocytosis.";  
RL Hum. Mutat. 8:97-107 (1996).  
RN [9]  
RP VARIANT EL2 SER-24.  
RX MEDLINE=94289716; PubMed=8018926;  
RA Parquet N., Devaux I., Boulanger L., Galand C., Boivin P.,  
RA Lecomte M.-C., Dhery D., Garbarz M.;  
RT "Identification of three novel spectrin alpha I/74 mutations in  
RT hereditary elliptocytosis: further support for a triple-stranded  
RT folding unit model of the spectrin heterodimer contact site.";  
RL Blood 84:303-308 (1994).  
RN [10]

```

spectrin Lyon (CTT-->TTT; alpha I 43 Leu-->Phe).";
RL Clin. Invest. 86:548-554(1990).
[18]
RN VARIANT EL2 JENDOUBA GLU-791.
RX MEDLINE=92345619; PubMed=1638030;
RA Alloisio N., Wilmette R., Morle L., Baklouti F., Marechal J.,
RA Ducluzeau M.-T., Denoroy L., Feo C., Forget B.G., Kastally R.,
RA Delanay J.;
RT "Spectrin Jendouba: an alpha II/31 spectrin variant that is
RT associated with elliptocytosis and carries a mutation distant from
RT the dimer self-association site.";
RL Blood 80:809-815(1992).
[19]
RN VARIANT EL2 TUNIS TRP-41.
RX MEDLINE=89323436; PubMed=2568861;
RA Morle L., Morie F., Roux A.F., Godet J., Forget B.G., Denoroy L.,
RA Garbarz M., Dhermy D., Kastally R., Delanay J.;
RT "Spectrin Tunis (Sp alpha I/78), an elliptocytogenic variant, is due
RT to the CGG->TCG codon change (Arg->Trp) at position 35 of the
RT alpha I domain.";
RL Blood 74:828-832(1989).
[20]
RN VARIANT EL2 GENOVA TRP-34.
RX MEDLINE=94250920; PubMed=8193371;
RA Perrotta S., del Giudice E.M., Alloisio N., Sciaratta G., Pinto L.,
RA Delanay J., Cuttillo S., Lolascen A.;
RT "Mild elliptocytosis associated with the alpha 34 Arg->Trp mutation
RT in spectrin Genova (alpha I/74).";
RL Blood 83:3346-3349(1994).
[21]
RN VARIANT EL2 ANASTASIA THR-45.
RX MEDLINE=95290423; PubMed=7772539;
RA Perrotta S., lolascen A., de Angelis F., Pagano L., Colonna G.,
RA Cuttillo S., del Giudice E.M.;
RT "Spectrin Anastasia (alpha I/78): a new spectrin variant (alpha 45
RT Arg->Thr) with moderate elliptocytogenic potential.";
RL Br. J. Haematol. 89:933-936(1995).
CC -!- FUNCTION: Spectrin is the major constituent of the cytoskeletal
CC network underlying the erythrocyte plasma membrane. It associates
CC with band 4.1 and actin to form the cytoskeletal superstructure of
CC the erythrocytic plasma membrane.
CC -!- SUBUNIT: Composed of nonhomologous chains, alpha and beta, which
CC aggregate side-to-side in an antiparallel fashion to form dimers,
CC tetramers, and higher polymers.
CC -!- DISEASE: Defects in SPPL are a cause of rheus-unlinked
CC elliptocytosis 2 (EL2) [MIM:130600, 182860, 166900].
CC Elliptocytosis (also known as ovalocytosis) is a genetically
CC heterogeneous, autosomal dominant hematologic disorder. It is
Query Match 3.1%; Score 159; DB 1; Length 2418;
Best Local Similarity 21.0%; Pred.No. 0.27;
Matches 131; Conservative 99; Mismatches 216; Indels 178; Gaps 33;
QY 79 ITSMDWTGKTAESLSLSDSKENNNTRISSMMSEVQRDNFYQHNVKL NVNSQL-SLDKS 137
DB 594 LKNWINKK-KLADDE-DYKD-----IQNLAKSVQKVFEK-ELAVNKTQTLENIQKT 643
QY 138 LTE-KSTOYLNOHQFAAMCK-----WONEKHTEQLLESEPQTVTLVE-----QFSNAN 186
DB 644 QGMIEGGHYASDNYVTTRLSEVASLN-----EELEATKKRGTKQLHEANOOLPEN-- 694
QY 187 IDRSPQDDHSDTSSENDNQOFLITVKLANAKOTTEDEHAREAKSHOKCSCHPED 246
DB 695 -----NAEDLORWLEDVEM-----QVTSEDIYGKLAEVQNLRKRKGLES 734
QY 247 CASCOQDEIDVPKPSPLSDVGVS---EDVGTGSKNDNKLIROESCIGNSPPEKESEP---- 300
DB 735 AVAAEQQVDI-----LTDLAAYPEIGHPDSDKIR-AQESELVCR---FEALKEPLATR 785
QY 301 -ESPMDVDNRSKNSCODSEADE-----ETSPGFDFQEDG-----SSSQTFANKPSRFQARDADI 351
DB 786 KKXLLDLLHLICRTDEEAEAWIOETSFATSYLGKLIASKGLINHRHVILENIASH 845

```

```

QY 352 EPRKRYSTKGEVRL---HFQEGGESRTGMDNLAKLPFNISLNVCECRNKKQHGKDS 408
DB 846 EPRIOETIERGNKMGVEGHFAEDVASR---VKSLNQ---NMESLFARARRQNDLEAV 899
QY 409 KITDHMLRLPKAEDRKEQWETHQSTERKIPKYVPPHSPDKMLGTPIEB-----MRKM 464
DB 900 QFOQYLADLHEATWIRE-----KEPTVDNTNYGADEEAAGALLKKHEAFLD 948
QY 465 PCGIELPLLRSANHTVIRVDLLRAGEVPKPPFTHYKLDNDKHVKMPCSEQNLYPVE 524
DB 949 NSFSGSMKALRNOAN-----ACQOQQAAPVE 974
QY 525 DENGERTAGSRWELIOTALLNKFTRPQNLKDALIKYNVASKKWDFTALI-----DFWDK 579
DB 975 GVAGEQRVWALYDF-----QARSREVT-----MKKGDVLTLLSSINKDWMK- 1016
QY 580 VLEBAEAQHYQSLPDMVKIALCLNICTQPIPLLKQKNH-----SITMSQBOI-----AS 632
DB 1017 -VEAAD-----HQGIIVP-----AVYVRLAHDEFPMLPQRRREBPNGNITQREQIENQYRS 1066
QY 633 LLANAFCTFPFRENKMKSEYSSY 656
DB 1067 LLDR-----EERRRLQRYNEF 1085

RESULT 13
ATFX DROME
ID ATFX DROME STANDARD; PRT; 1311 AA.
AC Q9GQN5; Q9VEQ8;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Transcriptional regulator ATFX homolog (x-linked nuclear protein)
DE (dXNP) (d-xnp).
GN XNP OR CG4548.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
[1]
RN SEQUENCE FROM N.A.
RP Cardoso C., Usseglio F., Villard L., Manfruell P., Rothbacher U.,
RX MEDLINE=20196006; PubMed=107311132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Ananides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Kandel M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballwey R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Fosler C., Gabrielian A.B., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D.A., Howland T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai X.,
RA Lasko P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,

```

```

RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Swigras R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195(2000).
[3]
RN SEQUENCE FROM N.A.
RP STRAIN=Berkley; TISSUE=Embryo;
RC MEDLINE=22426066; PubMed=12537569;
RX Stapleton M., Carlson J.W., Brokstein P., Yu C., Champe M.,
RA George R.A., Guarin H., Kronmiller B., Pacleb J.M., Park S., Wan K.H.,
RA Rubin G.M., Celniker S.E.;
RT "A Drosophila full-length cDNA resource."
RL Genome Biol. 3:RESEARCH0080.1-RESEARCH0080.8(2002).
CC -!- FUNCTION: Could be a global transcriptional regulator. Modifies
CC gene expression by affecting chromatin (Potential).
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- SIMILARITY: Belongs to the SNF2/RAD54 helicase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AF217802; AAG40586.1; -
CC EMBL; AB003752; AAF56471.1; -
CC EMBL; AY058592; AAL13821.1; -
CC FlyBase; FBgn0039338; XNP.
CC InterPro; IPR001410; DEAD.
CC InterPro; IPR001650; Helicase_C.
CC InterPro; IPR000330; SNF2_N.
CC Pfam; PF00271; helicase_C; 1.
CC Pfam; PF00176; SNF2_N; 1.
CC SMART; SM00487; DEXDC; 1.
CC SMART; SM00490; HELICG; 1.
KW DNA repair; Nuclear protein; DNA-binding; Helicase; ATP-binding.
FT NP_BIND 489 496 ATP (POTENTIAL).
FT SITE 615 618 DEGH BOX.
FT DOMAIN 69 74 POLY-SER.
FT DOMAIN 162 165 POLY-GLU.
FT DOMAIN 301 307 POLY-GLU.
FT CONFLICT 785 785 N -> T (IN REF. 1).
SQ SEQUENCE 1311 AA; 148218 MW; 71B78942468D4A68 CRC64;

Query Match 3.0%; Score 157.5; DB 1; Length 1311;
Best Local Similarity 18.0%; Pred. No. 0.15;
Matches 179; Conservative 142; Mismatches 372; Indels 299; Gaps 38;

QY 11 TKATRWGAATTPAASDARSFSPQRRLVDPDAHVQFVPPSPSPACVPCQACQHRGSAT 70
DB 25 SSVSRRESATESKSASESSESPPRS-----NTQSRTHKNVKGASGKATVSSSSDSQAVAN 80
QY 71 S-----LVFQKQITSMWMDTKGKTABESLDSKENNTNIESNMSSVQKD 116
DB 81 SSANDEKBPVKIRIVPLEKLAS-----PKTKERPGRGQKQKNTINDSDEEPLKG 134
QY 117 NFYQHNVEKLNVVNSQLSDK--SLTEKSTQYLNQHQHTAAMCKWQNEKGKTEQLLESEPOT 174
DB 135 S-----KLVLPAKSRKKNKASIIELS-----DSEVDEEESL 167

```

```
QY 175 VTLVPEQFNVANIDRSQNDHSDTDSENRDNOQFLTTVKLANA-----KQTTDE 226
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 168 LVAILPLP-KEAQOTKPKNSKASKSEIKRQKAQKEATTSSAIRASVNGTRGSLSE 226
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 227 HAREAKSHQKCS-----KSCPHGEDCASCQDQIDVWPKSPLSVDGSGEDVGTGSKNDNKLI 282
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 227 RSPASSRAESPFRPKCVLRKXVSLPKTKPAQPKQKSSD--SEEAATTSKSR--- 281
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 283 QBSCLNGSPPFKEKESPEPMVDNKNKSCQDSEADEETSFGDREDDGSSSTANKPS 342
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 282 -----QRRSKSEADSDYEPAPAAEEEEEERKSSG-DEBEAANSDDSYWMPQ 328
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 343 RFQAR-----DADIEFRKRYSTKGVEVLHFLQFEGESGRTGMDLNKLPGNISSLN 394
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 329 KRRRKSESQSGSDFEPEKEQKKGRK-RIKTTSGESGSDGDD----- 374
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 395 VECRNSKHGKDKSDITD---HLMLPKAEDRRKEQWETKHQ----- 433
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 375 -KQNKRGHRIKIITKDLDTTKEAKEEDRRKRIEDRQKLYNRFVKSESVEINELV 433
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 434 -----RTERKIPKVPVPHLSDPKKWLQ-----TPIEMRRMP-----RC-GI 469
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 434 LDFDEESKALLQVDKGLLKXPHQVAGVFKPMWDACFETILKESQKPKGSCILAHCMGL 493
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 470 RLPLLRPSANHTVTRVDLLRAGEVPKPFPTHYKDLWDNKHVK-MPCSEQNLVPVEDENG 528
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 494 GKTQVLTSLHTLVNTRTGVDRLVLIISPLSTVNNWARETSMWKFANRNDIEVYDI-- 551
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 529 ERTAGSRWELIQTALLNKFTPNQKQ-----AILKYNVAYSKKNDFTALIDFMDKVL 581
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 552 -----SRVK-----DKPTRIKLNWFNEGGVCILGYDM-----YRILANEKAKGL 592
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 582 BEAQAHIYQSIQIL---PDMVKIALCLPNICTQPIPLKQKNHSHITSQEQIASLLANAF 638
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 593 RKQOREQLMQALVDPGDLV-----VCDEGHLLKNEKTSISKAVT----- 632
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 639 FCTPRRNKMKSEVSSYPDINFNRLPEGRSRKPEKLTLCFVPRRYTEKTKTGLVFTT 698
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 633 -----RMTKRIVUTGTLQV-----NUREYVCMQFV--KPNLLGTYYK 670
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 699 QSLSDPPEWERCEKPLRLHVTYEGTIEENGQGMQVDFANRFV-----GGGVTSAGLVQ 754
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 671 -----YNNRFRVNPITNGQYTDSD--TE 690
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 755 BEIRFLINPELIISRLFTVLDHNECLIT-----GTEQYSEYTGATYR-----WSRSH 805
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 691 RDLRLMKHSHILHKLLEGCIQRDYSVLAPYLPKPKHYVYVYVTTLSLQOKLYGYMTTH 750
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 806 EDGSESD-----DCERRCTEIVADLHFRYRILDQFVPEKMRRELNKAYCGF 852
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 751 REQSGDVVGKARLFQDFQDLRRITWTH-----PMLRVNSDNDVIKALLNSDSDMEGF 805
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 853 LRPVGSSENLSAVATGNMGCGAFGGDARLKAL 884
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 806 ICDTDED--EAASNSSDSCETFKSDASMSGL 835
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
RESULT 14
ID BM2K MOUSE STANDARD; PRT: 1138 AA.
AC Q91256: OBC8L7;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DE 10-OCT-2003 (Rel. 42, Last annotation update)
DE BMP-2 inducible protein kinase (BC 2.7.1.37) (Bike).
GN BMP2K OR BIKE.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RC STRAIN=C57BL/6J;
```

```
RX MEDLINE=21551155; PubMed=11500515;
RA Kearns A.E., Donohue M.M., Sanyal B., Demay M.B.;
RT "cloning and characterization of a novel protein kinase that impairs
RL osteoblast differentiation in vitro.";
RN J. Biol. Chem. 276:42213-42218 (2001).
RP [2]
RC SEQUENCE FROM N.A. (ISOFORM 2).
RA STRAIN=C57BL/6J; TISSUE=Thymus;
RC MEDLINE=22354693; PubMed=12466851;
RA Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S.,
RA Nakado I., Osato N., Saito R., Suzuki H., Yamanaka I., Kiyosawa H.,
RA Yagi K., Tomaru Y., Hasegawa Y., Nogami A., Schonbach C., Gojobori T.,
RA Baldarelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J.,
RA Schirnel L.M., Kanapin A., Matsuda H., Batalov S., Keisel K.W.,
RA Blake J.A., Bradt D., Brusic V., Chothia C., Corbani L.E., Cousins S.,
RA Dalla E., Dragani T.A., Fletcher C.F., Forrest A., Frazer K.S.,
RA Gaasterland T., Gariboldi M., Gissi C., Godzik A., Gough J.,
RA Grimmond S., Gustincich S., Hirokawa N., Jackson I.J., Jarvis E.D.,
RA Kanai A., Kawaji H., Kawasawa Y., Kedzierski R.M., King B.L.,
RA Konagaya A., Kurochkin I.V., Lee Y., Lenhard B., Lyons P.A.,
RA Maglott D.R., Maltais L., Marchionni L., McKenzie L., Miki H.,
RA Nagasima T., Numata K., Okido T., Pavan W.J., Pertea G., Pesole G.,
RA Petrovsky N., Pillai R., Pontius J.U., Qi D., Ramachandran S.,
RA Ravasi T., Reed J.C., Reed D.J., Reid J., Ring B.Z., Ringwald M.,
RA Sandelin A., Schneider C., Semple C.A., Setou M., Shimada K.,
RA Sultana R., Takenaka Y., Taylor M.S., Teasdale R.D., Tomita M.,
RA Verardo R., Wagner L., Wahlestedt C., Wang Y., Watanabe Y., Wells C.,
RA Wilming L.G., Wyshaw-Boris A., Yanagisawa M., Yang I., Yang L.,
RA Yuan Z., Zavolan M., Zhu Y., Zimmer A., Carninci P., Hayatsu N.,
RA Hirozane-Kishikawa T., Konno H., Nakamura M., Sakazume N., Sato K.,
RA Shiraki T., Waki K., Kawai J., Aizawa K., Arakawa T., Fukuda S.,
RA Hara A., Hashizume W., Imotani K., Ishii Y., Itoh M., Kagawa I.,
RA Miyazaki A., Sakai K., Sakai D., Shibata K., Shinagawa A.,
RA Yaeunishi A., Yoshino M., Waterston R., Lander E.S., Rogers J.,
RA Birney E., Hayashizaki Y.;
RT "Analysis of the mouse transcriptome based on functional annotation of
RL 60,770 full-length cDNAs.";
RC Nature 420:563-573 (2002).
CC -!- FUNCTION: May be involved in osteoblast differentiation.
CC -!- CATALYTIC ACTIVITY: ATP + a protein = ADP + a phosphoprotein.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=1;
CC IsoId=Q91296-1; Sequence=Displayed;
CC Name=2;
CC IsoId=Q91296-2; Sequence=VSP_008094, VSP_008095;
CC -!- TISSUE SPECIFICITY: Expressed in spleen, kidney, lung, brain,
CC heart, diaphragm and calvaria but not in liver.
CC -!- PTM: Autophosphorylated.
CC -!- SIMILARITY: Belongs to the Ser/Thr family of protein kinases.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AY050249; AAC91585.1; -
CC EMBL; AK046752; BAC32854.1; -
CC PIR; PT0645; PT0645.
CC MGI; MGI-2155456; Bmp2k.
CC GO; GO:0005634; C:nucleus; IDA.
CC GO; GO:0019208; F:phosphatase regulator activity; IDA.
CC GO; GO:0004672; F:protein kinase activity; IDA.
CC GO; GO:0030500; P:regulation of bone mineralization; IDA.
CC InterPro; IPR000719; Prot Kinase.
CC InterPro; IPR008271; Ser_Thr_pkin_AS.
CC Pfam; PF00069; pkinase; I.
CC ProDom; PD000001; Prot_kinase; 1.
CC PROSITE; PS00107; PROTEIN_KINASE_ATP; FALSE_NEG.
```

DR PROSITE; PS00108; PROTEIN KINASE ST; 1.  
DR PROSITE; PS00111; PROTEIN KINASE DOM; 1.  
KW Transferase; Serine/threonine-protein kinase; Nuclear protein;  
KW ATE-binding; phosphorylation; Alternative splicing.  
FT DOMAIN 48 313 PROTEIN KINASE  
FT NP BIND 54 62 ATP (BY SIMILARITY).  
FT BINDING 76 77 ATP (BY SIMILARITY).  
FT ACT SITE 177 177 BY SIMILARITY.  
FT DOMAIN 421 546 GLN/HIS-RICH.  
FT VARSPLIC 638 649 NRGASTPSDKT -> SKGLKAYPASQ (in isoform 2).  
FT /FTID=VSP 008094.  
FT VARSPLIC 650 1138 Missing (in isoform 2).  
FT /FTID=VSP 008095.  
SQ SEQUENCE 1138 AA; 126185 MW; 8EDDE02F1333840D CRC64;  
Query Match 3.0%; Score 155; DB 1; Length 1138;  
Best Local Similarity 22.0%; Pred. No. 0.17;  
Matches 125; Conservative 69; Mismatches 201; Indels 174; Gaps 28;  
QY 54 SPACV-----PGQAG-----QHRGATSLVFK-----QKTIISWMDTKG----- 87  
Db 559 SPALVSYASSLPAGVGTIVDSYGNRSVAEKEAVANFTNQKTIHPDPMSCWNPFGEDN 618  
QY 88 -IKTAESSELDSEKNNNTRIESMGSVQKNFYQ-----HN---VEKLNVNSQLSLDKSL 138  
Db 619 FSKLTEEELD-REFDLNLSNRLGASTPSDKTVLPPAPHSRPPPEPFASVFFISHGSP 677  
QY 139 TEKSTQYL-NQHTAA-----MCKWQNEGKHT-----EQLLES 170  
Db 678 EKKTIHSPNQSITANLTNGGSSPLCKDQAGKTSNPVIRGVQVKGHDSDSDFS 737  
QY 171 EQPTVLVPEQFSNANIDRSPND--DHSDTDS-ENRDNQOFLITVKLANAKQTEDSH 227  
Db 738 DPPS-----PSSSEEEQEDDAQEGHGFNDDEPENLGRPLL-----WSEDEEDDKH 789  
QY 228 AREAKSHQCKSKCHPGEDCASCQDEIDVVPKSPISLVGSEVGVGSKNDKILQRESC 287  
Db 790 SSDSECEQAKTKR---GDTSLRLDKPGVAPDTALL-----TPARSPADALTSPQEPD 839  
QY 288 LNSGPPFEKESPEPMVDNKNKSCQSEAEETSPGDEQEDGSSSTANKPSRFQAR 347  
Db 840 VFGAVPFFAAPQSLQHRGDKNLSQA-----PFEQEDFVFTKA----- 881  
QY 348 DADIEPRKRYTKGGEVRLHFQEGGESRTGMNLMNAKLPGMISLNV-----ECR 398  
Db 882 ----PNNKVSQ-----DWPVAGPDR-----PLPARSRVDIFGSTPFQPFVS 923  
QY 399 NSQHGKKD-----SKITDHLMLPKAEDRRKEQWETHQKTERKIPKYVPHLSPDK 451  
Db 924 ASKSESKEDVFLVPPEITGS-QQQQKVQRSLQKLSSRQRTKQVSK-----SNKG 976  
QY 452 KWLGTPI-----EMRMRPRGIR-----LPLRPSANHTVIRVDLLR 490  
Db 977 RHGTPTSAKTLKPYRTPERARRHKVGRDRSQSSNEFLTISDSKENISVAITDGKDR 1036  
QY 491 AGVVPK-----PF---PHYKDLWNKH 510  
Db 1037 ASVLPDSDESLDPFGAKPFPPLDMLWQPH 1065  
RESULT 15  
SG1\_BOVIN  
ID SG1\_BOVIN STANDARD; PRT; 646 AA.  
AC P23389; 002707;  
DT 01-NOV-1991 (Rel. 20, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Secretogranin I precursor (Sgi) (Chromogranin B) (Cgb) [Contains: GAWK peptide; Secretolysin].  
GN CHGB.  
OS Bos taurus (Bovine).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
OC Bovidae; Bovinae; Bos.  
OX NCBI\_TaxID=9913;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Adrenal chromaffin;  
RX MEDLINE=91223091; PubMed=2025642;  
RA Bauer J.W., Fischer-Colbrie R.;  
RT "Primary structure of bovine chromogranin B deduced from cDNA sequence.";  
RL Biochim. Biophys. Acta 1089:124-126(1991).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Adrenal medulla;  
RX MEDLINE=97282588; PubMed=9136897;  
RA Yoo S.H., Kang Y.K.;  
RT "Identification of the secretory vesicle membrane binding region of chromogranin B.";  
RL FEBS Lett. 406:259-262(1997).  
RN [3]  
RP SEQUENCE OF 21-646 FROM N.A.  
RC TISSUE=Adrenal medulla;  
RA Grandy D.K., Leduc R., Makam H., Flanagan T., Diliberto E.J., Thomas G., Civelli O., Viveros O.H.;  
RL Submitted (Oct-1990) to the EMBL/GenBank/DBJ databases.  
RN [4]  
RP SEQUENCE OF 634-646.  
RC TISSUE=Adrenal chromaffin;  
RX MEDLINE=95262699; PubMed=7744058;  
RA Strub J.-M., Garcia-Sablone P., Lonnig K., Taupenot L., Hubert P., van Dorsselaer A., Anis D., Metz-Boutigue M.-H.;  
RT "Processing of chromogranin B in bovine adrenal medulla. Identification of secretolysin, the endogenous C-terminal fragment of residues 614-626 with antibacterial, the endogenous C-terminal activity.";  
RL Eur. J. Biochem. 229:356-368(1995).  
RN [5]  
RP CHARACTERIZATION OF SECRETOLYTIN.  
RX MEDLINE=96184581; PubMed=8603705;  
RA Strub J.M., Hubert P., Nullans G., Anis D., Metz-Boutigue M.-H.;  
RT "Antibacterial activity of secretolysin, a chromogranin B-derived peptide (614-626), is correlated with peptide structure.";  
RL FEBS Lett. 379:273-278(1996).  
CC -!- FUNCTION: Secretogranin I is a neuroendocrine secretory granule protein, which may be the precursor for other biologically active peptides. The 16 pairs of basic AA distributed throughout its sequence may be used as proteolytic cleavage sites.  
CC -!- FUNCTION: Secretolysin has antibacterial activity.  
CC -!- SUBCELLULAR LOCATION: Neuroendocrine and endocrine secretory granules.  
CC -!- PTM: O-glycosylated (Probable).  
CC -!- SIMILARITY: Belongs to the chromogranin / secretogranin protein family.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; X55027; CAA38846.1; -.  
DR EMBL; U88551; AAC48720.1; -.  
DR EMBL; X55489; CAA39109.1; -.  
DR PIR; S15901; S15901.  
DR InterPro; IPR001819; Chromogranin\_AB.  
DR InterPro; IPR001990; Granin.  
DR Pfam; PF01271; Granin; 1.  
DR PRINTS; PR00659; CHROMOGRANIN.  
DR PROSITE; PS00422; GRANINS 1; 1.  
DR PROSITE; PS00423; GRANINS 2; 1.  
KW Sulfation; Cleavage on pair of basic residues; Signal.  
FT SIGNAL 1 20

Search completed: May 26, 2004, 18:41:09  
Job time : 20.6372 secs

```

FT CHAIN          21      646      SECRETOGRANIN I.
FT PEPTIDE        418      484      GAWK PEPTIDE.
FT PEPTIDE        634      646      SECRETOLYTIN.
FT DISULFID       36       57      BY SIMILARITY.
FT MOD_RES        158      158      SULFATION (POTENTIAL).
FT MOD_RES        315      315      SULFATION (BY SIMILARITY).
FT CONFLICT       64       64      N -> S (IN REF. 1).
FT CONFLICT       70       70      N -> D (IN REF. 2).
FT CONFLICT       93       98      SEAPGL -> FRSPRAS (IN REF. 3).
FT CONFLICT      181      181      T -> M (IN REF. 2).
FT CONFLICT      261      261      H -> R (IN REF. 2).
FT CONFLICT      386      386      P -> R (IN REF. 2).
FT CONFLICT      481      481      H -> L (IN REF. 3).
FT CONFLICT      597      597      M -> V (IN REF. 2).
SQ SEQUENCE      646 AA; 73339 MW; 420DB1178FD9E415 CRC64;

Query Match          3.0%; Score 154.5; DB 1; Length 646;
Best Local Similarity 19.3%; Pred. No. 0.089;
Matches 144; Conservative 108; Mismatches 268; Indels 225; Gaps 36;

Qy 43 DAHVQFRVPPSPACVPGQAQGHRSATSLVFKOKTITSMMDTKGIK-----TAESLSLDS 98
Db 43 NALLKSSAPPITPEC--RQVLKKGKELKNEKSENENTREVRLLRDPADTSEAPGLSS 100

Qy 99 KENNNTRIESMMSSVQKDNFYQHNVEKLVNVSQSLDKSLTEKSTOYLNOHQTAACKWQ 158
Db 101 REDSGEG-DAQVPTVADTESGHSRERAGEPPGSAKEAKTRYKSEGGQNRREEMVKYQ 159

Qy 159 NEGK---HTEOLLESEP---QTVTIVPEQFSNANIDRSPQNDHSDTDSSENDRNQOFLT 212
Db 160 KREGEVGESEERLSEGPQKATFL-----NQRNQTAKKEELVSRDYDTQSAR 207

Qy 213 TVKLANAKQTTDEHARAKSHQKSC--HP-GEDCASQODEIDVVPKSPISDVGSE 269
Db 208 GLEKSHSRERSSQESGETKSOENWPQLQRHPEQGEAPG--ESEEDASPEV----- 257

Qy 270 DVGTSKNDNKLIRQESCLGNSPPEKESEPESPMDVDNSKNSCODSEADETSPGFDEQ 329
Db 258 -----DKRHSRPRHHGRSDRDRSQEGNPFL-----EESHVGV----- 291

Qy 330 EDGSSSTANKPSRPFQARDADIEFRKRYSTKGGEVRLHFQEGGHSRTGMNDLNAKLPGN 389
Db 292 TGNDSDEKARPAHFALEGAAY-----GEEVVRHSA-----AQAPGD 330

Qy 390 ISSLNVECRNKHQKQKSDKTDHMLRPLKADRRKQWETHKHQTERKIPKYVPHLSP 449
Db 331 LQGARFGGKRGGEH-----QALRRPSBES--LEQENKRH-----GLSP 366

Qy 450 DKKWLGTFIEM--REMPRCGIRLPLLR-----PSANHTVTIRVDLLR-AGEVPRKPFPTHY 502
Db 367 DLNMAQGYSESEBERGPAG--PSYRARGEEAAAYSTLGOTDEKRFGE-----THH 417

Qy 503 K---DLWDNKVMPKPCSEQNLYPVEDENGERTAGSRWELIQTALLNKFTRPQ-NLKDAIL 558
Db 418 RVQESQRDKARRRLPGELRNLYDYGEKGEAAARKWQ-----PQGDPRDA-- 463

Qy 559 KYNVAYKKWDFDALIDFWKVLBEAEQAHLVQSILPDMVKIALCLPNICTQPIPLLKOK 618
Db 464 -----DENREEARLRG-----KQY 477

Qy 619 MNHSITMSQEQIASLANAFF-----CTPPRR-----NAKMKSEYSSYPD 658
Db 478 APHHIT--EKRIGELL-NFFYDPSQWKSRSRFRKDPMDSDSFLGEEENGLTNEKNPFPE 534

Qy 659 INFN-----RLFE-----GRSSRKPEKLKTLFCYFRVRTEKKPTGLVTFTRQSLDFPWE 709
Db 535 YNYDWWKEKPFEDVNWGKKNPVKPLDKLQYDKRVAELDQ--LLHYRKKSAE-FPDFY 591

Qy 710 RCEKPLTRLHVTEGTIENGQGL 734
Db 592 DSEEQMSPOH-TAENEEXAGQGV 615
```



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 26, 2004, 17:19:25 ; Search time 34.3712 Seconds

(without alignments)  
2466.140 Million cell updates/sec

Title: US-09-302-812-4\_COPY\_677\_977

Perfect score: 1597

Sequence: 1 KTLFCYFRVVEKPTGLVT.....YHAVESCAETADHSQRTGT 300

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A Geneseq 29Jan04:\*

1: Geneseqp1980s:\*

2: Geneseqp1990s:\*

3: Geneseqp2000s:\*

4: Geneseqp2001s:\*

5: Geneseqp2002s:\*

6: Geneseqp2003as:\*

7: Geneseqp2003bs:\*

8: Geneseqp2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1597	100.0	976	5	Aae25630 Human pol
2	1597	100.0	976	5	Aau76021 Human pol
3	1597	100.0	976	5	Aau76012 Human pol
4	1597	100.0	976	6	Abg72279 Human pol
5	1497	93.7	968	5	Aae25631 Murine po
6	1497	93.7	968	5	Aau76022 Mouse pol
7	1497	93.7	968	5	Aau76013 Mouse pol
8	1497	93.7	968	6	Abg72280 Murine p
9	1493	93.5	977	5	Aae25629 Bovine po
10	1493	93.5	977	5	Aau76020 Bovine po
11	1493	93.5	977	5	Aau75799 Bovine po
12	1493	93.5	977	5	Abg72278 Bovine po
13	619.5	38.8	768	4	Abbs59491 Drosophil
14	619.5	38.8	768	5	Aae25632 Fruit fly
15	619.5	38.8	768	5	Aau76023 Fruit fly
16	619.5	38.8	768	5	Aau76014 Fruit fly
17	619.5	38.8	768	6	Abg72281 Fruit fly
18	371.5	23.3	726	5	Aae25633 Poly aden
19	371.5	23.3	726	5	Aau76024 Worm poly
20	371.5	23.3	726	5	Aau76015 Worm poly
21	371.5	23.3	726	6	Abg72282 C. elegan
22	344.5	21.6	819	4	Abg20721 Novel hum
23	210	13.1	100	4	Abg11103 Novel hum
24	178	11.1	33	5	Aae25652 Bovine po
25	167	10.5	31	5	Aae25634 Bovine po

## ALIGNMENTS

### RESULT 1

AAE25630

ID AAE25630 standard; protein; 976 AA.

XX

AC AAE25630;

XX 04-NOV-2002 (first entry)

DT Human poly adenosine diphosphate-ribose glycohydrolase (PARG).

XX Human; poly adenosine diphosphate-ribose glycohydrolase; PARG; enzyme;

XX ADP-ribose; neoplastic disorder; genotoxic oxidative stress; nontropic;

XX cardiac disorder; neuronal disorder; reperfusion injury; neurotoxicity;

XX Alzheimer's disease; Huntington's disease; Parkinson's disease; cardiant;

XX cytostatic; vasotropic; neuroprotective; anticonvulsant; gene therapy;

XX antisense therapy.

XX Homo sapiens.

XX US6395543-B1.

XX 28-MAY-2002.

XX 23-FEB-2000; 2000US-00511507.

XX 01-MAY-1998; 98US-0083768P.

XX 30-APR-1999; 99US-00302812.

XX (KENT ) UNIV KENTUCKY RES FOUND.

XX Jacobson MK, Jacobson EL, Ame J, Lin W;

XX WPI; 2002-535641/57.

XX N-PSDB; AAD42082.

XX New nucleic acid molecule encoding bovine poly adenosine diphosphate-

XX ribose glycohydrolase involved in cellular response to DNA damage,

XX inhibition of which is useful for treating neoplastic disorders and

XX neurodegenerative diseases.

XX Claim 3; Col 55-60; 77pp; English.

XX The invention relates to an isolated nucleic acid molecule which encodes

XX a bovine poly (adenosine diphosphate (ADP)-ribose) glycohydrolase (hPARG)

XX which catalyses release of ADP-ribose from an ADP-ribose polymer. The

XX invention is useful as probes and primer molecules that can used in

XX hybridisation assays and polymerase chain reaction (PCR) amplification.

XX The knowledge of the nucleotide sequence of the PARG gene permits the

Aae25651 Bovine po  
Aau76025 Bovine po  
Aau76016 Bovine po  
Abg72283 Oligopept  
Aae25635 Bovine po  
Aau76026 Bovine po  
Aau76017 Bovine po  
Abg72284 Oligopept  
Abu23332 Protein e  
Abg10064 Novel hum  
Abg18445 Novel hum  
Abg19877 Novel hum  
Abg14429 Novel hum  
Abg02294 Novel hum  
Abg10211 Novel hum  
Abg13269 Novel hum  
Abg10377 Novel hum  
Abg12499 Novel hum  
Abg07848 Novel hum  
Abg05427 Novel hum

26 167 10.5 31 5 AAE25651  
27 167 10.5 31 5 AAU76025  
28 167 10.5 31 5 AAU76016  
29 167 10.5 31 6 ABG72283  
30 149 9.3 29 5 AAE25635  
31 149 9.3 29 5 AAU76026  
32 149 9.3 29 5 AAU76017  
33 149 9.3 29 6 ABG72284  
34 97.5 6.1 395 6 ABU23332  
35 92.5 5.8 754 4 ABG10064  
36 92.5 5.8 754 4 ABG18445  
37 92.5 5.8 754 4 ABG19877  
38 92.5 5.8 754 4 ABG14429  
39 92.5 5.8 754 4 ABG02294  
40 92.5 5.8 754 4 ABG10211  
41 92.5 5.8 754 4 ABG13269  
42 92.5 5.8 754 4 ABG10377  
43 92.5 5.8 754 4 ABG12499  
44 92.5 5.8 754 4 ABG07848  
45 92.5 5.8 754 4 ABG05427

CC preparation of antisense therapeutics containing sequences complementary  
 CC to the mRNA of PARG gene. The antisense therapeutic are useful to treat  
 CC neoplastic disorders and conditions caused by genotoxic oxidative stress  
 CC e.g., cardiac disorders, neuronal disorders, reperfusion injury,  
 CC neurotoxicity, Alzheimer's disease, Huntington's disease and Parkinson's  
 CC disease. The invention is useful in gene therapy and antisense therapy.  
 CC The present sequence is human PARG  
 XX  
 XX  
 SQ Sequence 976 AA;  
 Query Match 100.0%; Score 1597; DB 5; Length 976;  
 Best Local Similarity 100.0%; Pred. No. 1.1e-166;  
 Matches 300; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 KTLFCYFRRVTEKPTGLVTFTRQSLDPPWERCCKPLRLHVTYEGTIEENGQGMLOV 60  
 DB 677 KTLFCYFRRVTEKPTGLVTFTRQSLDPPWERCCKPLRLHVTYEGTIEENGQGMLOV 736  
 QY 61 DFANRFVGGVTSAGLVQVEIRFLINPELIISRLFTVLDHNECLITGTEQYSEVTGYA 120  
 DB 737 DFANRFVGGVTSAGLVQVEIRFLINPELIISRLFTVLDHNECLITGTEQYSEVTGYA 796  
 QY 121 ETRWRSRSHEDGSRDDCERRCTEIVADALHFRYRLDQFVPEKMRRLNKAYCGFLRPG 180  
 DB 797 ETRWRSRSHEDGSRDDCERRCTEIVADALHFRYRLDQFVPEKMRRLNKAYCGFLRPG 856  
 QY 181 VSSNLSAVATGNWCGAFGGDARLKALIQILAAAAAERDVVYFTFGDSELMRDYISMHI 240  
 DB 857 VSSNLSAVATGNWCGAFGGDARLKALIQILAAAAAERDVVYFTFGDSELMRDYISMHI 916  
 QY 241 FLTERKLTGVDVYKLLRYNNEECRCNCTPGPDIKLYPFYIHAVESCAETADHSGQRTGT 300  
 DB 917 FLTERKLTGVDVYKLLRYNNEECRCNCTPGPDIKLYPFYIHAVESCAETADHSGQRTGT 976  
 RESULT 2  
 AAU76021  
 ID AAU76021 standard; protein; 976 AA.  
 AC AAU76021;  
 DT 08-MAY-2002 (first entry)  
 DE Human poly(ADP-ribose) glycohydrolase (PARG) protein sequence.  
 KW Human; poly(ADP-ribose) glycohydrolase; PARG; DNA repair; ADP-ribose;  
 KW adenosine diphosphate-ribose; neoplastic disorder; adenocarcinoma;  
 KW leukaemia; lymphoma; melanoma; myeloma; sarcoma; teratocarcinoma;  
 KW hyperplasia; hypertrophy; reperfusion; ischaemia; heart attack; stroke;  
 KW neurodegenerative disease; neurological disorder; Alzheimer's disease;  
 KW Huntington's disease; Parkinson's disease.  
 XX  
 OS Homo sapiens.  
 XX  
 XX US6337202-B1.  
 XX  
 XX 08-JAN-2002.  
 XX  
 XX 23-FEB-2000; 2000US-00511477.  
 XX  
 XX 01-MAY-1998; 98US-0083768P.  
 XX  
 XX 30-APR-1999; 99US-00302812.  
 XX  
 XX (KENT ) UNIV KENTUCKY RES FOUND.  
 XX  
 XX Jacobson MK, Jacobson EL, Ame J, Lin W;  
 XX WPI; 2002-163240/21.  
 XX  
 XX N-PSDB; ABK14932.  
 XX  
 XX Novel isolated and purified poly (ADP-ribose) glycohydrolase protein  
 XX PT which catalyses release of ADP-ribose from ADP ribose polymer, useful for  
 XX PT treating neoplastic and neurological disorders, heart attack and stroke.

XX  
 XX  
 XX  
 CC Claim 2; Col 55-60; 81pp; English.  
 CC  
 CC The present invention relates to a new poly(ADP-ribose) glycohydrolase  
 CC (PARG) protein which catalyses release of ADP-ribose from an ADP  
 CC (adenosine diphosphate)-ribose polymer. The PARG molecule of the  
 CC invention is useful for generating antibodies and can be inhibited or  
 CC activated for diagnosing and treating neoplastic disorders such as  
 CC adenocarcinoma, leukaemia, lymphoma, melanoma, myeloma, sarcoma,  
 CC teratocarcinoma, hyperplasia and hypertrophy, reperfusion following  
 CC ischaemia, heart attack, stroke, neurodegenerative diseases, neurological  
 CC disorders including Alzheimer's, Huntington's and Parkinson's diseases,  
 CC and related conditions. PARG levels may be enhanced to suppress DNA  
 CC repair and increase the cell's susceptibility to chemotherapy drugs.  
 CC Antagonists of PARG are administered to treat or prevent neoplastic  
 CC disorders. The present amino acid sequence represents the human PARG  
 CC protein of the invention. This protein is one of several PARG proteins  
 CC (AAU76020-AAU76024) of the invention  
 XX  
 XX  
 SQ Sequence 976 AA;  
 Query Match 100.0%; Score 1597; DB 5; Length 976;  
 Best Local Similarity 100.0%; Pred. No. 1.1e-166;  
 Matches 300; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 KTLFCYFRRVTEKPTGLVTFTRQSLDPPWERCCKPLRLHVTYEGTIEENGQGMLOV 60  
 DB 677 KTLFCYFRRVTEKPTGLVTFTRQSLDPPWERCCKPLRLHVTYEGTIEENGQGMLOV 736  
 QY 61 DFANRFVGGVTSAGLVQVEIRFLINPELIISRLFTVLDHNECLITGTEQYSEVTGYA 120  
 DB 737 DFANRFVGGVTSAGLVQVEIRFLINPELIISRLFTVLDHNECLITGTEQYSEVTGYA 796  
 QY 121 ETRWRSRSHEDGSRDDCERRCTEIVADALHFRYRLDQFVPEKMRRLNKAYCGFLRPG 180  
 DB 797 ETRWRSRSHEDGSRDDCERRCTEIVADALHFRYRLDQFVPEKMRRLNKAYCGFLRPG 856  
 QY 181 VSSNLSAVATGNWCGAFGGDARLKALIQILAAAAAERDVVYFTFGDSELMRDYISMHI 240  
 DB 857 VSSNLSAVATGNWCGAFGGDARLKALIQILAAAAAERDVVYFTFGDSELMRDYISMHI 916  
 QY 241 FLTERKLTGVDVYKLLRYNNEECRCNCTPGPDIKLYPFYIHAVESCAETADHSGQRTGT 300  
 DB 917 FLTERKLTGVDVYKLLRYNNEECRCNCTPGPDIKLYPFYIHAVESCAETADHSGQRTGT 976  
 RESULT 3  
 AAU76012  
 ID AAU76012 standard; protein; 976 AA.  
 AC AAU76012;  
 XX  
 XX 08-MAY-2002 (first entry)  
 DT  
 DE Human poly(ADP-ribose) glycohydrolase (PARG) protein sequence.  
 XX  
 XX Human; poly(ADP-ribose) glycohydrolase; PARG; PARG modulator; ADP-ribose;  
 XX adenosine diphosphate-ribose; DNA repair; apoptosis; neoplasia; cancer;  
 XX inherited genetic disease; myocardial infarction; vascular stroke; aging;  
 XX neurodegeneration; Huntington's disease; Parkinson's disease;  
 XX Alzheimer's disease; neurotoxicity.  
 XX  
 OS Homo sapiens.  
 XX  
 XX US6333148-B1.  
 XX  
 XX 25-DEC-2001.  
 XX  
 XX 30-APR-1999; 99US-00302812.  
 XX  
 XX 01-MAY-1998; 98US-0083768P.  
 XX  
 XX (KENT ) UNIV KENTUCKY RES FOUND.

XX Jacobson MK, Jacobson EL, Ame J, Lin W;  
 XX WPI; 2002-153820/20.  
 XX N-PSDB; ABK14494.  
 XX Screening compounds for modulation of poly(ADP-ribose) glycohydrolase,  
 PT useful potentially for treating diseases associated with DNA damage, e.g.  
 PT cancer.  
 XX Claim 3; Col 55-60; 80pp; English.  
 XX The present invention relates to a new method for screening compounds for  
 CC ability to modulate activity of an enzyme that hydrolyses ADP (adenosine  
 CC diphosphate)-ribose from an ADP-ribose polymer. The compounds are  
 CC inhibitors or activators of PARG (poly(ADP-ribose) glycohydrolase) and  
 CC are used to treat or prevent any condition associated with DNA damage  
 CC e.g. neoplasia, inherited genetic diseases, myocardial infarction,  
 CC vascular stroke, aging and neurodegeneration e.g. Huntington's,  
 CC Parkinson's or Alzheimer's diseases, or neurotoxicity generally.  
 CC Compounds identified by the new method are more effective than known  
 CC inhibitors and have fewer side effects. The present amino acid sequence  
 CC represents the human PARG protein of the invention. This protein is one  
 CC of several PARG proteins (AAU75799 and AAU76012-AAU76015) of the  
 CC invention  
 XX SQ Sequence 976 AA;  
 Query Match 100.0%; Score 1597; DB 5; Length 976;  
 Best Local Similarity 100.0%; Pred. No. 1.1e-166;  
 Matches 300; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 KTLFCYFRVTEKPTGLVTFTRQSLDPPWEKCEKPLTRLHVYEGTIEENGQMLQV 60  
 Db 677 KTLFCYFRVTEKPTGLVTFTRQSLDPPWEKCEKPLTRLHVYEGTIEENGQMLQV 736  
 QY 61 DFANRFVGGVTSAGLVQVEIRFLINPELIIISRLFTEVLDHNECLIIITGQYSEYTGVA 120  
 Db 737 DFANRFVGGVTSAGLVQVEIRFLINPELIIISRLFTEVLDHNECLIIITGQYSEYTGVA 796  
 QY 121 ETYRWSRSHEDGSDRDCERRCTEIVADALHFRYLDQFVPEKMRRELKAYCGFLRPG 180  
 Db 797 ETYRWSRSHEDGSDRDCERRCTEIVADALHFRYLDQFVPEKMRRELKAYCGFLRPG 856  
 QY 181 VSSENLSAVATGNWCGAFGGDARLKALIQIIILAAAAARDVYFTFGDSELMRDYISMHI 240  
 Db 857 VSSENLSAVATGNWCGAFGGDARLKALIQIIILAAAAARDVYFTFGDSELMRDYISMHI 916  
 QY 241 FLTERKLTVDGVYKLLRYNEECRNCSPTGPDIKLYPIYHAVESCAETADHSGORTGT 300  
 Db 917 FLTERKLTVDGVYKLLRYNEECRNCSPTGPDIKLYPIYHAVESCAETADHSGORTGT 976  
 RESULT 4  
 ABG72279  
 ID ABG72279 standard; protein; 976 AA.  
 XX ABG72279;  
 XX 13-MAR-2003 (first entry)  
 XX Human poly(ADP-ribose) glycohydrolase (PARG) enzyme.  
 XX Human; poly(ADP-ribose) glycohydrolase; PARG; PARG expression;  
 KW cellular response; DNA damage; neoplastic disorder inducing agent;  
 KW oxidative stress; neoplastic disorder; myocardial infarction;  
 KW vascular stroke; neurodegenerative disorder; Alzheimer's disease;  
 KW Parkinson's disease; Huntington's disease; inborn genetic error;  
 KW reperfusion; ischaemia; aging; neurotoxicity; PARG activity; cytostatic;  
 KW neuroprotective; nootropic; antiparkinsonian; cardiant; vasotropic;  
 KW anticonvulsant; cerebroprotective; enzyme.  
 XX Homo sapiens.  
 OS

XX US2002132328-A1.  
 XX 19-SEP-2002.  
 XX 09-OCT-2001; 2001US-00973451.  
 XX 01-MAY-1998; 98US-0083768P.  
 XX 30-APR-1999; 99US-00302812.  
 XX (JACO/) JACOBSON M K.  
 XX (JACO/) JACOBSON E L.  
 XX (AMEJ/) AME J.  
 XX (LINW/) LIN W.  
 XX Jacobson MK, Jacobson EL, Ame J, Lin W;  
 XX WPI; 2003-155895/15.  
 XX N-PSDB; ABK14478.  
 XX New nucleic acid molecule encoding a polypeptide with poly(ADP-ribose)  
 PT glycohydrolase activity, for preventing, treating, or ameliorating a  
 PT disease condition, e.g. neoplastic disorder, myocardial infarction or  
 PT vascular stroke.  
 XX Claim 28; Fig 16; 86pp; English.  
 XX The present invention relates to the isolation of poly(ADP-ribose)  
 CC glycohydrolase (PARG) from several species, and the polynucleotide  
 CC sequences encoding them. Methods for inhibiting PARG expression or  
 CC overexpressing PARG are also disclosed. PARG is involved in the cellular  
 CC response to DNA damage, and is associated with the body's response to  
 CC neoplastic disorder inducing agents and oxidative stress. The  
 CC polynucleotide sequences encoding PARG and PARG modulators are useful for  
 CC preventing, treating, or ameliorating diseases such as neoplastic  
 CC disorders, myocardial infarction, vascular stroke, neurodegenerative  
 CC diseases (e.g. Alzheimer's disease, Parkinson's disease or Huntington's  
 CC disease), inborn genetic errors, reperfusion following ischaemia, aging,  
 CC and neurotoxicity. The polynucleotide sequences are also useful in gene  
 CC therapy. The methods are useful for identifying an agent that modulates  
 CC PARG activity, identifying a mutant PARG allele in an individual, or  
 CC screening candidate molecules for PARG modulating activity. The present  
 CC sequence represents human PARG enzyme  
 XX SQ Sequence 976 AA;  
 Query Match 100.0%; Score 1597; DB 6; Length 976;  
 Best Local Similarity 100.0%; Pred. No. 1.1e-166;  
 Matches 300; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 KTLFCYFRVTEKPTGLVTFTRQSLDPPWEKCEKPLTRLHVYEGTIEENGQMLQV 60  
 Db 677 KTLFCYFRVTEKPTGLVTFTRQSLDPPWEKCEKPLTRLHVYEGTIEENGQMLQV 736  
 QY 61 DFANRFVGGVTSAGLVQVEIRFLINPELIIISRLFTEVLDHNECLIIITGQYSEYTGVA 120  
 Db 737 DFANRFVGGVTSAGLVQVEIRFLINPELIIISRLFTEVLDHNECLIIITGQYSEYTGVA 796  
 QY 121 ETYRWSRSHEDGSDRDCERRCTEIVADALHFRYLDQFVPEKMRRELKAYCGFLRPG 180  
 Db 797 ETYRWSRSHEDGSDRDCERRCTEIVADALHFRYLDQFVPEKMRRELKAYCGFLRPG 856  
 QY 181 VSSENLSAVATGNWCGAFGGDARLKALIQIIILAAAAARDVYFTFGDSELMRDYISMHI 240  
 Db 857 VSSENLSAVATGNWCGAFGGDARLKALIQIIILAAAAARDVYFTFGDSELMRDYISMHI 916  
 QY 241 FLTERKLTVDGVYKLLRYNEECRNCSPTGPDIKLYPIYHAVESCAETADHSGORTGT 300  
 Db 917 FLTERKLTVDGVYKLLRYNEECRNCSPTGPDIKLYPIYHAVESCAETADHSGORTGT 976  
 RESULT 5  
 AAE25631

ID XX AAE25631 standard; protein; 968 AA.  
 AC XX AAE25631;  
 XX DT 04-NOV-2002 (first entry)  
 XX DE Murine poly adenosine diphosphate-ribose glycohydrolase (PARG).  
 XX DE Murine; poly adenosine diphosphate-ribose glycohydrolase; PARG; enzyme;  
 KW ADP-ribose; neoplastic disorder; genotoxic oxidative stress; neurotropic;  
 KW cardiac disorder; neuronal disorder; reperfusion injury; neurotoxicity;  
 KW Alzheimer's disease; Huntington's disease; Parkinson's disease; cardiant;  
 KW cytosatic; vasotropic; neuroprotective; anticonvulsant; gene therapy;  
 KW antisense therapy.  
 XX OS Mus musculus.  
 XX PN US6395543-B1.  
 XX PD 28-MAY-2002.  
 XX PF 23-FEB-2000; 2000US-00511507.  
 XX PR 01-MAY-1998; 98US-0083769P.  
 XX PR 30-APR-1999; 99US-00302812.  
 XX PA (KENT ) UNIV KENTUCKY RES FOUND.  
 XX PI Jacobson MK, Jacobson EL, Ame J, Lin W;  
 XX WPI; 2002-535641/57.  
 XX N-PSDB; RAD42083.  
 XX PT New nucleic acid molecule encoding bovine poly adenosine diphosphate-  
 PT ribose glycohydrolase involved in cellular response to DNA damage,  
 PT inhibition of which is useful for treating neoplastic disorders and  
 PT neurodegenerative diseases.  
 XX PS Claim 3; Col 63-68; 77pp; English.  
 XX CC The invention relates to an isolated nucleic acid molecule which encodes  
 CC a bovine poly (adenosine diphosphate (ADP)-ribose) glycohydrolase (bPARG)  
 CC which catalyses release of ADP-ribose from an ADP ribose polymer. The  
 CC invention is useful as probes and primer molecules that can be used in  
 CC hybridisation assays and polymerase chain reaction (PCR) amplification.  
 CC The knowledge of the nucleotide sequence of the PARG gene permits the  
 CC preparation of antisense therapeutics containing sequences complementary  
 CC to the mRNA of PARG gene. The antisense therapeutic are useful to treat  
 CC neoplastic disorders and conditions caused by genotoxic oxidative stress  
 CC e.g., cardiac disorders, neuronal disorders, reperfusion injury,  
 CC neurotoxicity, Alzheimer's disease, Huntington's disease and Parkinson's  
 CC disease. The invention is useful in gene therapy and antisense therapy.  
 XX Sequence 968 AA;  
 Query Match 93.7%; Score 1497; DB 5; Length 968;  
 Best Local Similarity 93.7%; Pred. No. 1.3e-155;  
 Matches 281; Conservative 7; Mismatches 12; Indels 0; Gaps 0;  
 QY 1 KTLFCYFRVTEKTKPTGLVTFTRQSLDFFPERCEKPLTRLHVYEGTIENGGMQLQV 60  
 DB 669 KTLFCYFRVTEKTKPTGLVTFTRQSLDFFPERCEKPLTRLHVYEGTIENGGMQLQV 728  
 QY 61 DFANRFVGGVTSAGVQVEIRFLINPELIIIRLFTFVLDRNECLIIITGEQSYTGVA 120  
 DB 729 DFANRFVGGVTSAGVQVEIRFLINPELIIIRLFTFVLDRNECLIIITGEQSYTGVA 788  
 QY 121 EYRWSRSHEDGSDRDCRRCTEIVADLHFRYLDQFVPEKRELNKAYCCFLRPG 180  
 DB 789 EYRWSRSHEDGSDRDCRRCTEIVADLHFRYLDQFVPEKRELNKAYCCFLRPG 848  
 QY 181 VSENLSAVATGNWGCAGFGDARLKALIQILAAAAARDVVYFTFGDSELMRDIYSMHT 240

DB 849 VSENLSAVATGNWGCAGFGDARLKALIQILAAAAARDVVYFTFGDSELMRDIYSMHT 908  
 QY 241 FLTERKLTWGDVYKLLRLRYNNECRNCSTPGPDIKLYPFIYHVESCAETADHSGQRTGT 300  
 DB 909 FLTERKLDVGKVKLLRLRYNNECRNCSTPGPDIKLYPFIYHVESCAETADHSGQRTGT 968  
 RESULT 6  
 AAU76022  
 ID AAU76022 standard; protein; 968 AA.  
 XX AC AAU76022;  
 XX DT 08-MAY-2002 (first entry)  
 XX DE Mouse poly(ADP-ribose) glycohydrolase (PARG) protein sequence.  
 XX KW Mouse; poly(ADP-ribose) glycohydrolase; PARG; DNA repair; ADP-ribose;  
 KW adenosine diphosphate-ribose; neoplastic disorder; adenocarcinoma;  
 KW leukaemia; lymphoma; melanoma; myeloma; sarcoma; teratocarcinoma;  
 KW hyperplasia; hypertrophy; reperfusion; ischaemia; heart attack; stroke;  
 KW neurodegenerative disease; neurological disorder; Alzheimer's disease;  
 KW Huntington's disease; Parkinson's disease.  
 XX OS Mus musculus.  
 XX PN US6337202-B1.  
 XX PD 08-JAN-2002.  
 XX PF 23-FEB-2000; 2000US-00511477.  
 XX PR 01-MAY-1998; 98US-0083768P.  
 XX PR 30-APR-1999; 99US-00302812.  
 XX PA (KENT ) UNIV KENTUCKY RES FOUND.  
 XX PI Jacobson MK, Jacobson EL, Ame J, Lin W;  
 XX WPI; 2002-163240/21.  
 XX N-PSDB; ABK14933.  
 XX PT Novel isolated and purified poly (ADP-ribose) glycohydrolase protein  
 PT which catalyses release of ADP-ribose from ADP ribose polymer, useful for  
 PT treating neoplastic and neurological disorders, heart attack and stroke.  
 XX PS Claim 2; Col 63-70; 81pp; English.  
 XX CC The present invention relates to a new poly(ADP-ribose) glycohydrolase  
 CC (PARG) protein which catalyses release of ADP-ribose from an ADP  
 CC (adenosine diphosphate)-ribose polymer. The PARG molecule of the  
 CC invention is useful for generating antibodies and can be inhibited or  
 CC activated for diagnosing and treating neoplastic disorders such as  
 CC adenocarcinoma, leukaemia, lymphoma, melanoma, myeloma, sarcoma,  
 CC teratocarcinoma, hyperplasia and hypertrophy, reperfusion following  
 CC ischaemia, heart attack, stroke, neurodegenerative diseases, neurological  
 CC disorders including Alzheimer's, Huntington's and Parkinson's diseases,  
 CC and related conditions. PARG levels may be enhanced to suppress DNA  
 CC repair and increase the cell's susceptibility to chemotherapy drugs.  
 CC Antagonists of PARG are administered to treat or prevent neoplastic  
 CC disorders. The present amino acid sequence represents the mouse PARG  
 CC protein of the invention. This protein is one of several PARG proteins  
 CC (AAU76020-AAU76024) of the invention  
 XX Sequence 968 AA;  
 Query Match 93.7%; Score 1497; DB 5; Length 968;  
 Best Local Similarity 93.7%; Pred. No. 1.3e-155;  
 Matches 281; Conservative 7; Mismatches 12; Indels 0; Gaps 0;  
 QY 1 KTLFCYFRVTEKTKPTGLVTFTRQSLDFFPERCEKPLTRLHVYEGTIENGGMQLQV 60

Db 669 KTLFCYFRVTEKPTGLVTFTRQSLDEPWEKCEKELTRHLVTEGTEIENGKGMLOV 728  
QY 61 DPNARFVGGVTSAGLVQVEIRFLINPELIIISRLFTVLDHNECLIIITGTEQSYSTGYA 120  
Db 729 DPNARFVGGVTSAGLVQVEIRFLINPELIIISRLFTVLDHNECLIIITGTEQSYSTGYA 788  
QY 121 ETYRWSRSHEDGSEKDDCERRCTEIVADALHFRYLDQFVPEKVRRELKAYCGFLRPG 180  
Db 789 ETYRWSRSHEDGSEKDDCERRCTEIVADALHFRYLDQFVPEKVRRELKAYCGFLRPG 848  
QY 181 VSENLISAVATGNWCGAFGGDARLKALIQILAAAAERDVVYFTFGDSELMDIYSMH 240  
Db 849 VSENLISAVATGNWCGAFGGDARLKALIQILAAAAERDVVYFTFGDSELMDIYSMH 908  
QY 241 FLTERKLTVDGVYKLLRYNEECNCSTPGDIIKLYPIYHVESCAETADHSQORTGT 300  
Db 909 FLTERKLTVDGVYKLLRYNEECNCSTPGDIIKLYPIYHVESCAETADHSQORTGT 968

RESULT 7

AAU76013  
ID AAU76013 standard; protein; 968 AA.  
AC AAU76013;

XX 08-MAY-2002 (first entry)

XX Mouse poly(ADP-ribose) glycohydrolase (PARG) protein sequence.

XX Mouse; poly(ADP-ribose) glycohydrolase; PARG; PARG modulator; ADP-ribose;  
XX adenosine diphosphate-ribose; DNA repair; apoptosis; neoplasia; cancer;  
XX inherited genetic disease; myocardial infarction; vascular stroke; aging;  
XX neurodegeneration; Huntington's disease; Parkinson's disease;  
XX Alzheimer's disease; neurotoxicity.

XX Mus musculus.

XX US6333148-B1.

XX 25-DEC-2001.

XX 30-APR-1999; 99US-00302812.

XX 01-MAY-1998; 98US-0083768P.

XX (KENT ) UNIV KENTUCKY RES FOUND.

XX Jacobson MK, Jacobson EL, Ame J, Lin W;

XX WPI; 2002-153820/20.

XX N-PSDB; ABK14495.

XX Screening compounds for modulation of poly(ADP-ribose) glycohydrolase,  
XX useful potentially for treating diseases associated with DNA damage, e.g.  
XX cancer.

XX Claim 3; Col 63-68; 80pp; English.

XX The present invention relates to a new method for screening compounds for  
XX ability to modulate activity of an enzyme that hydrolyses ADP (adenosine  
XX diphosphate)-ribose from an ADP-ribose polymer. The compounds are  
XX inhibitors or activators of PARG (poly(ADP-ribose) glycohydrolase) and  
XX are used to treat or prevent any condition associated with DNA damage,  
XX e.g. neoplasia, inherited genetic diseases, myocardial infarction,  
XX vascular stroke, aging and neurodegeneration e.g. Huntington's,  
XX Parkinson's or Alzheimer's diseases, or neurotoxicity generally.  
XX Compounds identified by the new method are more effective than known  
XX inhibitors and have fewer side effects. The present amino acid sequence  
XX represents the mouse PARG protein of the invention. This protein is one  
XX of several PARG proteins (AAU75799 and AAU76012-AAU76015) of the  
XX invention

XX Sequence 968 AA;

Query Match 93.7%; Score 1497; DB 5; Length 968;

Best Local Similarity 93.7%; Pred. No. 1.3e-155;

Matches 281; Conservative 7; Mismatches 12; Indels 0; Gaps 0;

QY 1 KTLFCYFRVTEKPTGLVTFTRQSLDEPWEKCEKELTRHLVTEGTEIENGKGMLOV 60  
Db 669 KTLFCYFRVTEKPTGLVTFTRQSLDEPWEKCEKELTRHLVTEGTEIENGKGMLOV 728  
QY 61 DPNARFVGGVTSAGLVQVEIRFLINPELIIISRLFTVLDHNECLIIITGTEQSYSTGYA 120  
Db 729 DPNARFVGGVTSAGLVQVEIRFLINPELIIISRLFTVLDHNECLIIITGTEQSYSTGYA 788  
QY 121 ETYRWSRSHEDGSEKDDCERRCTEIVADALHFRYLDQFVPEKVRRELKAYCGFLRPG 180  
Db 789 ETYRWSRSHEDGSEKDDCERRCTEIVADALHFRYLDQFVPEKVRRELKAYCGFLRPG 848  
QY 181 VSENLISAVATGNWCGAFGGDARLKALIQILAAAAERDVVYFTFGDSELMDIYSMH 240  
Db 849 VSENLISAVATGNWCGAFGGDARLKALIQILAAAAERDVVYFTFGDSELMDIYSMH 908  
QY 241 FLTERKLTVDGVYKLLRYNEECNCSTPGDIIKLYPIYHVESCAETADHSQORTGT 300  
Db 909 FLTERKLTVDGVYKLLRYNEECNCSTPGDIIKLYPIYHVESCAETADHSQORTGT 968

RESULT 8

ABG72280  
ID ABG72280 standard; protein; 968 AA.

XX AC ABG72280;

XX DT 13-MAR-2003 (first entry)

XX Mmurine poly(ADP-ribose) glycohydrolase (PARG) enzyme.

XX Murine; poly(ADP-ribose) glycohydrolase; PARG; PARG expression;  
XX cellular response; DNA damage; neoplastic disorder inducing agent;  
XX oxidative stress; neoplastic disorder; myocardial infarction;  
XX vascular stroke; neurodegenerative disease; Alzheimer's disease;  
XX Parkinson's disease; Huntington's disease; inborn genetic error;  
XX reperfusion; ischaemia; aging; neurotoxicity; PARG activity; mouse;  
XX cytosolic; neuroprotective; nontropic; antiparkinsonian; cardiac;  
XX vasotropic; anticonvulsant; cerebroprotective; enzyme.

XX Mus musculus.

XX US2002132328-A1.

XX 19-SEP-2002.

XX 09-OCT-2001; 2001US-00973451.

XX 01-MAY-1998; 98US-0083768P.

XX 30-APR-1999; 99US-00302812.

XX (JACO/) JACOBSON M K.

XX (JACO/) JACOBSON E L.

XX (AMEJ/) AME J.

XX (LINW/) LIN W.

XX Jacobson MK, Jacobson EL, Ame J, Lin W;

XX WPI; 2003-155895/15.

XX N-PSDB; ABK14479.

XX New nucleic acid molecule encoding a polypeptide with poly(ADP-ribose)  
XX glycohydrolase activity, for preventing, treating, or ameliorating a  
XX disease condition, e.g. neoplastic disorder, myocardial infarction or  
XX vascular stroke.

XX Claim 28; Fig 16; 86pp; English.

CC The present invention relates to the isolation of poly(ADP-ribose)  
 CC glycohydrolase (PARG) from several species, and the polynucleotide  
 CC sequences encoding them. Methods for inhibiting PARG expression or  
 CC overexpressing PARG are also disclosed. PARG is involved in the cellular  
 CC response to DNA damage, and is associated with the body's response to  
 CC polynucleotide disorder inducing agents and oxidative stress. The  
 CC polynucleotide sequences encoding PARG and PARG modulators are useful for  
 CC preventing, treating, or ameliorating diseases such as neoplastic  
 CC disorders, myocardial infarction, vascular stroke, neurodegenerative  
 CC disorders (e.g. Alzheimer's disease, Parkinson's disease or Huntington's  
 CC disease), inborn genetic errors, reperfusion following ischemia, aging,  
 CC and neurotoxicity. The polynucleotide sequences are also useful in gene  
 CC therapy. The methods are useful for identifying an agent that modulates  
 CC PARG activity, identifying a mutant PARG allele in an individual, or  
 CC screening candidate molecules for PARG modulating activity. The present  
 CC sequence represents murine PARG enzyme  
 CC  
 CC SQ

Query Match 93.7%; Score 1497; DB 6; Length 968;  
 Best Local Similarity 93.7%; Pred. No. 1.3e-155;  
 Matches 281; Conservative 7; Mismatches 12; Indels 0; Gaps 0;  
 QY 1 KTLFCYFRVTEKPTGLVTFTRQSLDFPWEKCEKPLRLHVTYEGTIEGNGQMLQV 60  
 DB 669 KTLFCYFRVTEKPTGLVTFTRQSLDFPWEKCEKPLRLHVTYEGTIEGNGQMLQV 728  
 QY 61 DFNRFVGGVTSAGLQVEIRFLINPELIIISRLFTFVLDHNECLITGTQYSEYTGVA 120  
 DB 729 DFNRFVGGVTSAGLQVEIRFLINPELIIISRLFTFVLDHNECLITGTQYSEYTGVA 788  
 QY 121 ETYRWSRSHEDGSDRCERCTEIVADLHFRYLDQFVPEKMRRLNKAYCGFLRPG 180  
 DB 789 ETYRWSRSHEDGSDRCERCTEIVADLHFRYLDQFVPEKMRRLNKAYCGFLRPG 848  
 QY 181 VSENLSAVATGNWCGGAFGGDARLKALIQILAAAAAERDVVYTFGDSLMRDIYSMHI 240  
 DB 849 VSENLSAVATGNWCGGAFGGDARLKALIQILAAAAAERDVVYTFGDSLMRDIYSMHI 908  
 QY 241 FLTERKLTGVGVYKLLRYNNECRNCSTPGPDIKLYPFIYHVESCAETADHSQRTG 300  
 DB 909 FLTERKLTGVGVYKLLRYNNECRNCSTPGPDIKLYPFIYHVESCAETADHSQRTG 969

RESULT 9  
 AAE25629  
 ID AAE25629 standard; protein; 977 AA.  
 AC AAE25629;  
 DT 04-NOV-2002 (first entry)  
 XX Bovine poly adenosine diphosphate-ribose glycohydrolase (bPARG).  
 DE  
 KW Bovine; poly adenosine diphosphate-ribose glycohydrolase; bPARG; enzyme;  
 KW ADP-ribose; neoplastic disorder; genotoxic oxidative stress; neurotoxic;  
 KW cardiac disorder; neuronal disorder; reperfusion injury; neurotoxicity;  
 KW Alzheimer's disease; Huntington's disease; Parkinson's disease; cardiac;  
 KW cytotostatic; vasotropic; neuroprotective; anticonvulsant; gene therapy;  
 KW antisense therapy.  
 XX  
 OS Bos taurus.  
 XX  
 PN US6395543-B1.  
 XX  
 PD 28-MAY-2002.  
 XX  
 PF 23-FEB-2000; 2000US-00511507.  
 XX  
 PR 01-MAY-1998; 98US-0083768P.  
 PR 30-APR-1999; 99US-00302812.  
 XX  
 PA (KENT ) UNIV KENTUCKY RES FOUND.

XX Jacobson MK, Jacobson EL, Ame J, Lin W;  
 XX WPI; 2002-535641/57.  
 XX N-PSDB; AAD42081.  
 XX  
 XX New nucleic acid molecule encoding bovine poly adenosine diphosphate-  
 XX ribose glycohydrolase involved in cellular response to DNA damage,  
 XX PT inhibition of which is useful for treating neoplastic disorders and  
 XX neurodegenerative diseases.  
 XX  
 XX Claim 3; Col 47-45; 77pp; English.  
 XX  
 XX The invention relates to an isolated nucleic acid molecule which encodes  
 XX a bovine poly (adenosine diphosphate (ADP)-ribose) glycohydrolase (bPARG)  
 XX which catalyzes release of ADP-ribose from an ADP ribose polymer. The  
 XX invention is useful as probes and primer molecules that can be used in  
 XX hybridisation assays and polymerase chain reaction (PCR) amplification.  
 XX The knowledge of the nucleotide sequence of the PARG gene permits the  
 XX preparation of antisense therapeutics containing sequences complementary  
 XX to the mRNA of PARG gene. The antisense therapeutics are useful to treat  
 XX neoplastic disorders and conditions caused by genotoxic oxidative stress  
 XX e.g. cardiac disorders, neuronal disorders, reperfusion injury,  
 XX e.g. cardiac disorders, neuronal disorders, reperfusion injury,  
 XX neurotoxicity, Alzheimer's disease, Huntington's disease and Parkinson's  
 XX disease. The invention is useful in gene therapy and antisense therapy.  
 XX The present sequence is bPARG  
 XX  
 XX SQ

Query Match 93.5%; Score 1493; DB 5; Length 977;  
 Best Local Similarity 94.0%; Pred. No. 3.6e-155;  
 Matches 281; Conservative 7; Mismatches 11; Indels 0; Gaps 0;  
 QY 1 KTLFCYFRVTEKPTGLVTFTRQSLDFPWEKCEKPLRLHVTYEGTIEGNGQMLQV 60  
 DB 678 KTLFCYFRVTEKPTGLVTFTRQSLDFPWEKCEKPLRLHVTYEGTIEGNGQMLQV 737  
 QY 61 DFNRFVGGVTSAGLQVEIRFLINPELIIISRLFTFVLDHNECLITGTQYSEYTGVA 120  
 DB 738 DFNRFVGGVTSAGLQVEIRFLINPELIIISRLFTFVLDHNECLITGTQYSEYTGVA 797  
 QY 121 ETYRWSRSHEDGSDRCERCTEIVADLHFRYLDQFVPEKMRRLNKAYCGFLRPG 180  
 DB 798 ETYRWSRSHEDGSDRCERCTEIVADLHFRYLDQFVPEKMRRLNKAYCGFLRPG 857  
 QY 181 VSENLSAVATGNWCGGAFGGDARLKALIQILAAAAAERDVVYTFGDSLMRDIYSMHI 240  
 DB 858 VSENLSAVATGNWCGGAFGGDARLKALIQILAAAAAERDVVYTFGDSLMRDIYSMHI 917  
 QY 241 FLTERKLTGVGVYKLLRYNNECRNCSTPGPDIKLYPFIYHVESCAETADHSQRTG 299  
 DB 918 FLTERKLTGVGVYKLLRYNNECRNCSTPGPDIKLYPFIYHVESCAETADHSQRTG 976

RESULT 10  
 AAU76020  
 ID AAU76020 standard; protein; 977 AA.  
 XX  
 AC AAU76020;  
 XX  
 DT 08-MAY-2002 (first entry)  
 XX  
 XX Bovine poly (ADP-ribose) glycohydrolase (PARG) protein sequence.  
 DE  
 KW Cow; poly (ADP-ribose) glycohydrolase; PARG; DNA repair; ADP-ribose;  
 KW adenosine diphosphate-ribose; neoplastic disorder; adenocarcinoma;  
 KW leukaemia; lymphoma; melanoma; myeloma; sarcoma; teratocarcinoma;  
 KW hyperplasia; hypertrophy; reperfusion; ischaemia; heart attack; stroke;  
 KW neurodegenerative disease; neurological disorder; Alzheimer's disease;  
 KW Huntington's disease; Parkinson's disease.  
 XX  
 OS Bos taurus.

Key Location/Qualifiers  
601. .617  
/note= "Represents PARG oligopeptide #1"  
Region  
761. .770  
/note= "Represents PARG oligopeptide #2"  
Region  
771. .801  
/note= "Represents PARG oligopeptide #3"  
Region  
849. .880  
/note= "Represents PARG oligopeptide #4"  
US6337202-B1.  
08-JAN-2002.  
23-FEB-2000; 2000US-00511477.  
01-MAY-1998; 98US-0083768P.  
30-APR-1999; 99US-00302812.  
(KENT ) UNIV KENTUCKY RES FOUND.  
Jacobson MK, Jacobson EL, Ame J, Lin W;  
WPI: 2002-163240/21.  
N-PSDB; ABK14931.  
Novel isolated and purified poly (ADP-ribose) glycohydrolase protein  
which catalyses release of ADP-ribose from ADP ribose polymer, useful for  
treating neoplastic and neurological disorders, heart attack and stroke.  
Claim 2; Col 47-52; 81pp; English.  
The present invention relates to a new poly(ADP-ribose) glycohydrolase  
(PARG) protein which catalyses release of ADP-ribose from an ADP  
(adenosine diphosphate)-ribose polymer. The PARG molecule of the  
invention is useful for generating antibodies and can be inhibited or  
activated for diagnosing and treating neoplastic disorders such as  
adenocarcinoma, leukaemia, lymphoma, melanoma, myeloma, sarcoma,  
teratocarcinoma, hyperplasia and hypertrophy, reperfusion following  
ischaemia, heart attack, stroke, neurodegenerative diseases, neurological  
disorders including Alzheimer's, Huntington's and Parkinson's diseases,  
and related conditions. PARG levels may be enhanced to suppress DNA  
repair and increase the cell's susceptibility to chemotherapy drugs.  
Antagonists of PARG are administered to treat or prevent neoplastic  
disorders. The present amino acid sequence represents the bovine PARG  
protein of the invention. This protein is one of several PARG proteins  
(AAU76020-AAU76024) of the invention  
Sequence 977 AA;  
Query Match 93.5%; Score 1493; DB 5; Length 977;  
Best Local Similarity 94.0%; Pred. No. 3.6e-155;  
Matches 281; Conservative 7; Mismatches 11; Indels 0; Gaps 0;  
QY 1 KTLFCYFRVTEKKPTGLVTFTRQSLDEPWEKCEKPLTRLHVTEGTEENGGQMLQV 60  
Db 678 KTLFCYFRVTEKKPTGLVTFTRQSLDEPWEKCEKPLTRLHVTEGTEENGGQMLQV 737  
QY 61 DFANRVGGVTSAGLVQVEIRFLINPELIIISRLFTVEILDHNECLITGTQYSEYTGVA 120  
Db 738 DFANRVGGVTSAGLVQVEIRFLINPELIIISRLFTVEILDHNECLITGTQYSEYTGVA 797  
QY 121 ETYRWSRSHEDGSRDDCERRCTTEI VALDALHFRYLDQFVPEKRELNKAYCGFLRPG 180  
Db 798 ETYRWSRSHEDGSRDDCERRCTTEI VALDALHFRYLDQFVPEKRELNKAYCGFLRPG 857  
QY 181 VSSENLSAVATGNGCGAFGGDARLKALIQILAAARDDVYVFTGDSLEMDIYSMHII 240  
Db 858 VSSENLSAVATGNGCGAFGGDARLKALIQILAAARDDVYVFTGDSLEMDIYSMHII 917  
QY 241 FLTERKLTVGDVYKLLIRYNEECNCSCTPGDIPKLYPFYHYAVESCAETADHSQRTG 299  
Db 918 FLTERKLTVGEVYKLLIRYNEECNCSCTPGDIPKLYPFYHYAVESCTQTNPQGRGTG 976

RESULT 11  
AAU75799  
ID AAU75799 standard; protein; 977 AA.  
XX AAU75799;  
XX 08-MAY-2002 (first entry)  
XX Bovine poly(ADP-ribose) glycohydrolase (PARG) protein sequence.  
XX Cow; poly(ADP-ribose) glycohydrolase; PARG; PARG modulator; ADP-ribose;  
XX adenosine diphosphate-ribose; DNA repair; apoptosis; neoplasia; cancer;  
XX inherited genetic disease; myocardial infarction; vascular stroke; aging;  
XX neurodegeneration; Huntington's disease; Parkinson's disease;  
XX Alzheimer's disease; neurotoxicity.  
XX Bos taurus.  
XX Key Location/Qualifiers  
XX Region 601. .617  
XX FT /note= "Represents PARG oligopeptide #1"  
XX Region 761. .770  
XX FT /note= "Represents PARG oligopeptide #2"  
XX Region 771. .801  
XX FT /note= "Represents PARG oligopeptide #3"  
XX Region 849. .880  
XX FT /note= "Represents PARG oligopeptide #4"  
XX US6333148-B1.  
XX 25-DEC-2001.  
XX 30-APR-1999; 99US-00302812.  
XX 01-MAY-1998; 98US-0083768P.  
XX (KENT ) UNIV KENTUCKY RES FOUND.  
XX Jacobson MK, Jacobson EL, Ame J, Lin W;  
XX WPI: 2002-153820/20.  
XX N-PSDB; ABK14493.  
XX Screening compounds for modulation of poly(ADP-ribose) glycohydrolase,  
XX useful potentially for treating diseases associated with DNA damage, e.g.  
XX cancer.  
XX Claim 3; Col 45-52; 80pp; English.  
XX The present invention relates to a new method for screening compounds for  
XX ability to modulate activity of an enzyme that hydrolyses ADP (adenosine  
XX diphosphate)-ribose from an ADP-ribose polymer. The compounds are  
XX inhibitors or activators of PARG (poly(ADP-ribose) glycohydrolase) and  
XX are used to treat or prevent any condition associated with DNA damage  
XX e.g. neoplasia, inherited genetic diseases, myocardial infarction,  
XX vascular stroke, aging and neurodegeneration e.g. Huntington's,  
XX Parkinson's or Alzheimer's diseases, or neurotoxicity generally.  
XX Compounds identified by the new method are more effective than known  
XX inhibitors and have fewer side effects. The present amino acid sequence  
XX represents the bovine PARG protein of the invention. This protein is one  
XX of several PARG proteins (AAU75799 and AAU76012-AAU76015) of the  
XX invention  
XX Sequence 977 AA;  
Query Match 93.5%; Score 1493; DB 5; Length 977;  
Best Local Similarity 94.0%; Pred. No. 3.6e-155;  
Matches 281; Conservative 7; Mismatches 11; Indels 0; Gaps 0;  
QY 1 KTLFCYFRVTEKKPTGLVTFTRQSLDEPWEKCEKPLTRLHVTEGTEENGGQMLQV 60

678 KTLFCYFRRVTEKKPTGLVTFTRQSLDPEWERCCKLRLHVTVEGTIEGNGQMLQV 737  
 61 DFANRFVGGVTSAGLVQVEIRFLINPELIIISRLFTVLDHNECLITGTQYSEYTGVA 120  
 738 DFANRFVGGVTSAGLVQVEIRFLINPELIIISRLFTVLDHNECLITGTQYSEYTGVA 797  
 121 ETYRWSRSHEDGSDRCCTEIVAIIDALHFRFVLDQVPEKMRRLKAYCGFLRPG 180  
 798 ETYRWSRSHEDGSDRCCTEIVAIIDALHFRFVLDQVPEKMRRLKAYCGFLRPG 857  
 181 VSSENISAVATGNWCGAFGGDARLKALIQILAAAAAERDVVYFTFGDSLMRDIYSMHI 240  
 858 VSSENISAVATGNWCGAFGGDARLKALIQILAAAAAERDVVYFTFGDSLMRDIYSMHI 917  
 241 FLTERKLTGVDVYKLLRYNEECNCTGPDIKLYPIYHVESCAETADHSGQRTG 299  
 918 FLTERKLTGVDVYKLLRYNEECNCTGPDIKLYPIYHVESCTQTNPQGRGTG 976

RESULT 12  
 ABG72278  
 ID ABG72278 standard; protein; 977 AA.  
 XX  
 AC ABG72278;  
 XX  
 DT 13-MAR-2003 (first entry)  
 XX  
 DE Bovine poly (ADP-ribose) glycohydrolase (PARG) enzyme.  
 XX  
 KW Bovine; poly (ADP-ribose) glycohydrolase; PARG; PARG expression;  
 KW cellular response; DNA damage; neoplastic disorder inducing agent;  
 KW oxidative stress; neoplastic disorder; myocardial infarction;  
 KW vascular stroke; neurodegenerative disorder; Alzheimer's disease;  
 KW Parkinson's disease; Huntington's disease; inborn genetic error;  
 KW reperfusion; ischaemia; aging; neurotoxicity; PARG activity; cytostatic;  
 KW neuroprotective; nootropic; antiparkinsonian; cardiac; vasotropic;  
 KW anticonvulsant; cerebroprotective; enzyme.  
 XX  
 OS Bos taurus.  
 XX  
 PN US2002132328-A1.  
 XX  
 PD 19-SEP-2002.  
 XX  
 PF 09-OCT-2001; 2001US-00973451.  
 XX  
 PR 01-MAY-1998; 98US-0083768P.  
 PR 30-APR-1999; 99US-00302812.  
 XX  
 PA (JACO/) JACOBSON M K.  
 PA (JACO/) JACOBSON E L.  
 PA (AMEJ/) AME J.  
 PA (LINW/) LIN W.  
 XX  
 PI Jacobson MK, Jacobson EL, Ame J, Lin W;  
 XX  
 DR WPI; 2003-155895/15.  
 DR N-PSDB; ABX14477.  
 XX  
 PT New nucleic acid molecule encoding a polypeptide with poly(ADP-ribose)  
 PT glycohydrolase activity, for preventing, treating, or ameliorating a  
 PT disease condition, e.g. neoplastic disorder, myocardial infarction or  
 PT vascular stroke.  
 XX  
 PS Claim 28; Fig 16; 86pp; English.  
 XX  
 CC The present invention relates to the isolation of poly(ADP-ribose)  
 CC glycohydrolase (PARG) from several species, and the polynucleotide  
 CC sequences encoding them. Methods for inhibiting PARG expression or  
 CC overexpressing PARG are also disclosed. PARG is involved in the cellular  
 CC response to DNA damage, and is associated with the body's response to  
 CC neoplastic disorder inducing agents and oxidative stress. The  
 CC polynucleotide sequences encoding PARG and PARG modulators are useful for

CC preventing, treating, or ameliorating diseases such as neoplastic  
 CC disorders, myocardial infarction, vascular stroke, neurodegenerative  
 CC disorders (e.g. Alzheimer's disease, Parkinson's disease or Huntington's  
 CC disease), inborn genetic errors, reperfusion following ischaemia, aging,  
 CC and neurotoxicity. The polynucleotide sequences are also useful in gene  
 CC therapy. The methods are useful for identifying an agent that modulates  
 CC PARG activity, identifying a mutant PARG allele in an individual, or  
 CC screening candidate molecules for PARG modulating activity. The present  
 CC sequence represents bovine PARG enzyme  
 XX

Query Match 93.5%; Score 1493; DB 6; Length 977;  
 Best Local Similarity 94.0%; Pred. No. 3.6e-155;  
 Matches 281; Conservative 7; Mismatches 11; Indels 0; Gaps 0;  
 QY 1 KTLFCYFRRVTEKKPTGLVTFTRQSLDPEWERCCKLRLHVTVEGTIEGNGQMLQV 60  
 DB 678 KTLFCYFRRVTEKKPTGLVTFTRQSLDPEWERCCKLRLHVTVEGTIEGNGQMLQV 737  
 QY 61 DFANRFVGGVTSAGLVQVEIRFLINPELIIISRLFTVLDHNECLITGTQYSEYTGVA 120  
 DB 738 DFANRFVGGVTSAGLVQVEIRFLINPELIIISRLFTVLDHNECLITGTQYSEYTGVA 797  
 QY 121 ETYRWSRSHEDGSDRCCTEIVAIIDALHFRFVLDQVPEKMRRLKAYCGFLRPG 180  
 DB 798 ETYRWSRSHEDGSDRCCTEIVAIIDALHFRFVLDQVPEKMRRLKAYCGFLRPG 857  
 QY 181 VSSENISAVATGNWCGAFGGDARLKALIQILAAAAAERDVVYFTFGDSLMRDIYSMHI 240  
 DB 858 VSSENISAVATGNWCGAFGGDARLKALIQILAAAAAERDVVYFTFGDSLMRDIYSMHI 917  
 QY 241 FLTERKLTGVDVYKLLRYNEECNCTGPDIKLYPIYHVESCAETADHSGQRTG 299  
 DB 918 FLTERKLTGVDVYKLLRYNEECNCTGPDIKLYPIYHVESCTQTNPQGRGTG 976

RESULT 13  
 ABB59491  
 ID ABB59491 standard; protein; 768 AA.  
 XX  
 AC ABB59491;  
 XX  
 DT 26-MAR-2002 (first entry)  
 XX  
 DE Drosophila melanogaster polypeptide SEQ ID NO 5265.  
 XX  
 KW Drosophila; developmental biology; cell signalling; insecticide;  
 KW pharmaceutical.  
 XX  
 OS Drosophila melanogaster.  
 XX  
 PN WO200171042-A2.  
 XX  
 PD 27-SEP-2001.  
 XX  
 PF 23-MAR-2001; 2001WO-US009231.  
 XX  
 PR 23-MAR-2000; 2000US-0191637P.  
 PR 11-JUL-2000; 2000US-00614150.  
 XX  
 PA (PEKE) PE CORP NY.  
 XX  
 PI Venter JC, Adams M, Li PWD, Myers EW;  
 XX  
 DR WPI; 2001-656860/75.  
 DR N-PSDB; ABL03594.  
 XX  
 PT New isolated nucleic acid detection reagent for detecting 1000 or more  
 PT genes from Drosophila and for elucidating cell signaling and cell-cell  
 PT interactions.  
 XX  
 PS Disclosure; SEQ ID NO 5265; 21pp + Sequence Listing; English.





Job time : 36.3712 secs

XX OS Drosophila melanogaster.  
XX PN US6337202-B1.  
XX XX  
XX PD 08-JAN-2002.  
XX XX  
XX XX 23-FEB-2000; 2000US-00511477.  
XX XX  
XX XX 01-MAY-1998; 98US-0083768P.  
XX PR 30-APR-1999; 99US-00302812.  
XX XX  
XX XX (KENT ) UNIV KENTUCKY RES FOUND.  
XX PA  
XX XX Jacobson MK, Jacobson EL, Ame J, Lin W;  
XX XX  
XX XX WPI; 2002-163240/21.  
XX DR N-PSDB; ABK14934.  
XX XX  
XX PT Novel isolated and purified poly (ADP-ribose) glycohydrolase protein  
XX PT which catalyses release of ADP-ribose from ADP ribose polymer, useful for  
XX PT treating neoplastic and neurological disorders, heart attack and stroke.  
XX XX  
XX PS Example 13; Col 71-76; 81pp; English.  
XX XX  
XX CC The present invention relates to a new poly(ADP-ribose) glycohydrolase  
XX CC (PARG) protein which catalyses release of ADP-ribose from an ADP  
XX CC (adenosine diphosphate)-ribose polymer. The PARG molecule of the  
XX CC invention is useful for generating antibodies and can be inhibited or  
XX CC activated for diagnosing and treating neoplastic disorders such as  
XX CC adenocarcinoma, leukaemia, lymphoma, melanoma, myeloma, sarcoma,  
XX CC teratocarcinoma, hyperplasia and hypertrophy, reperfusion following  
XX CC ischaemia, heart attack, stroke, neurodegenerative diseases, neurological  
XX CC disorders including Alzheimer's, Huntington's and Parkinson's diseases,  
XX CC and related conditions. PARG levels may be enhanced to suppress DNA  
XX CC repair and increase the cell's susceptibility to chemotherapy drugs.  
XX CC Antagonists of PARG are administered to treat or prevent neoplastic  
XX CC disorders. The present amino acid sequence represents the fruit fly PARG  
XX CC protein of the invention. This protein is one of several PARG proteins  
XX CC (AAU76020-AAU76024) of the invention  
XX SQ  
SQ Sequence 768 AA;  
Query Match 38.8%; Score 619.5; DB 5; Length 768;  
Best Local Similarity 46.4%; Pred. NO. 98-59;  
Matches 140; Conservative 41; Mismatches 90; Indels 31; Gaps 8;  
QY 1 KTLFCYFRV--TEKK---PTGLVTFTRQS--LEDFPEWERCERPL--TRLHVTYEGTI 50  
Db 297 KCMHYFRVCTERDASNVPTGVTFVRSGLPEHLIDWSQSAAPLGDVPLHVDAGETI 356  
QY 51 BENGQMLQVDFANRFVGGVTSAGLVQBEIRFLINPELIIIRLFTVLDHNECLITGT 110  
Db 357 EDEGIGLLQVDFANKYLGGLGVGHGCVQBEIRFVICPELLVKGKLFTECLRPFALVMLGA 416  
QY 111 EYSEYTGVAETVRNRSRSHEDSRRDDCERRCTEIVAIIDALHRRYLDQFVPEKMERELN 170  
Db 417 ERYSNYTGAGSEWNGNFEDSTPDSGRRQTAVAIIDALHFAQSHHQTRDLMERELN 476  
QY 171 KAYCGFLR-----PGVSSENLSAVATGNWCGAFGGDARLKALIQILAAAAERDVVYF 224  
Db 477 KAYIGFVHMVTPFG-----VATGNWCGAFGGDSYKALLQMLVCAQLGRPLAYY 528  
QY 225 TFGDSLMRDYISMHIFLTERKLTVDGVYKLLRYNEECRNCSTPGP-----DIKLYP 278  
Db 529 TFGNVEFRDDFHEMWLLFRNDGTTVQLWS-ILRSYSRLIKEKSKREPRENKASKKKLYD 587  
QY 279 FI 280  
Db 588 FI 589

Search completed: May 26, 2004, 18:40:03

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 26, 2004, 18:38:05 ; Search time 9.61306 Seconds  
(without alignments)  
1611.120 Million cell updates/sec

Title: US-09-302-812-4\_COPY\_677\_977

Perfect score: 1597

Sequence: 1 KTLFCYFRRVTEKKPTGLVT.....YHAVESCAETHSGQRTGT 300

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA:\*

1: /cgn2\_6/ptodata/2/iaa/5A\_COMB.pep.\*

2: /cgn2\_6/ptodata/2/iaa/5B\_COMB.pep.\*

3: /cgn2\_6/ptodata/2/iaa/6A\_COMB.pep.\*

4: /cgn2\_6/ptodata/2/iaa/6B\_COMB.pep.\*

5: /cgn2\_6/ptodata/2/iaa/PCTUS\_COMB.pep.\*

6: /cgn2\_6/ptodata/2/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1597	100.0	976	4	US-09-302-812-4
2	1597	100.0	976	4	US-09-511-477-4
3	1597	100.0	976	4	US-09-511-507-4
4	1497	93.7	968	4	US-09-302-812-6
5	1497	93.7	968	4	US-09-511-477-6
6	1497	93.7	968	4	US-09-511-507-6
7	1493	93.5	977	4	US-09-302-812-2
8	1493	93.5	977	4	US-09-511-477-2
9	1493	93.5	977	4	US-09-511-507-2
10	619.5	38.8	768	4	US-09-302-812-8
11	619.5	38.8	768	4	US-09-511-477-8
12	619.5	38.8	768	4	US-09-511-507-8
13	371.5	23.3	726	4	US-09-302-812-10
14	371.5	23.3	726	4	US-09-511-477-10
15	371.5	23.3	726	4	US-09-511-507-10
16	167	10.5	31	4	US-09-302-812-11
17	167	10.5	31	4	US-09-511-477-11
18	167	10.5	31	4	US-09-511-507-11
19	149	9.3	29	4	US-09-302-812-12
20	149	9.3	29	4	US-09-511-477-12
21	149	9.3	29	4	US-09-511-507-12
22	91.5	5.7	261	2	US-08-691-814B-4
23	90	5.6	749	4	US-09-328-352-7588
24	89	5.6	205	4	US-08-630-915A-16
25	85	5.3	355	4	US-09-489-039A-8367
26	85	5.3	580	3	US-09-234-393-48
27	85	5.3	580	4	US-09-865-171-48

28	85	5.3	581	3	US-09-234-393-20	Sequence 20, Appl
29	85	5.3	581	3	US-09-234-393-44	Sequence 44, Appl
30	85	5.3	581	3	US-09-234-393-46	Sequence 46, Appl
31	85	5.3	581	4	US-09-865-171-20	Sequence 20, Appl
32	85	5.3	581	4	US-09-865-171-44	Sequence 44, Appl
33	85	5.3	581	4	US-09-865-171-46	Sequence 46, Appl
34	85	5.3	581	4	US-09-398-395A-48	Sequence 48, Appl
35	85	5.3	581	4	US-09-887-586A-48	Sequence 48, Appl
36	85	5.3	581	4	US-09-895-752-48	Sequence 48, Appl
37	85	5.3	581	4	US-09-903-012B-48	Sequence 48, Appl
38	85	5.3	581	4	US-09-900-797-48	Sequence 48, Appl
39	83.5	5.2	505	4	US-09-620-405B-478	Sequence 478, App
40	83.5	5.2	505	4	US-09-620-405B-485	Sequence 485, App
41	83.5	5.2	505	4	US-09-604-287A-478	Sequence 478, App
42	83.5	5.2	505	4	US-09-604-287A-485	Sequence 485, App
43	83.5	5.2	505	4	US-09-834-759-478	Sequence 478, App
44	83.5	5.2	505	4	US-09-834-759-485	Sequence 485, App
45	83.5	5.2	915	2	US-08-480-917-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1

US-09-302-812-4  
; Sequence 4, Application US/09302812B  
; Patent No. 6333148  
; GENERAL INFORMATION:  
; APPLICANT: JACOBSON, Myron K.  
; APPLICANT: JACOBSON, Elaine L.  
; APPLICANT: AM, Jean-Christophe  
; APPLICANT: LIN, Winston  
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOPOLYMERASE (PARG) ENZ  
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIVE  
; TITLE OF INVENTION: THEREWITH  
; FILE REFERENCE: NIAID 201  
; CURRENT APPLICATION NUMBER: US/09/302,812B  
; CURRENT FILING DATE: 1999-04-30  
; EARLIER APPLICATION NUMBER: 60/083,768  
; EARLIER FILING DATE: 1998-05-01  
; NUMBER OF SEQ ID NOS: 38  
; SEQ ID NO 4  
; LENGTH: 976  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
US-09-302-812-4

Query Match	100.0%;	Score 1597;	DB 4;	Length 976;
Best Local Similarity	100.0%;	Pred. No. 2.1e-171;		
Matches	300;	Conservative	0;	Mismatches 0; Indels 0; Gaps 0;
QY	1	KTLFCYFRRVTEKKPTGLVTFTTQSLDPPPEWRCCKPLRLHVTVTGTTEENGQGMLOV	60	
Db	677	KTLFCYFRRVTEKKPTGLVTFTTQSLDPPPEWRCCKPLRLHVTVTGTTEENGQGMLOV	736	
QY	61	DFANRFVGGVTSAGLVQBEIRFLINPELLISLFTVLDHNECLITGTQYSEYTGVA	120	
Db	737	DFANRFVGGVTSAGLVQBEIRFLINPELLISLFTVLDHNECLITGTQYSEYTGVA	796	
QY	121	ETVWRSRSHDGERDDCERCCTEIVADALHFRFRLDQFVPEKMRRLNKAYCGFLRPG	180	
Db	797	ETVWRSRSHDGERDDCERCCTEIVADALHFRFRLDQFVPEKMRRLNKAYCGFLRPG	856	
QY	181	VSSENLSAVATGNWCGAFGCGDARLKALIQILAAAAARDVVFTFGDSLMRDIDYSMHI	240	
Db	857	VSSENLSAVATGNWCGAFGCGDARLKALIQILAAAAARDVVFTFGDSLMRDIDYSMHI	916	
QY	241	FLTERKLTVDGVVKLLRLRYNNEECRCNCTPGDPDKLYPFIYHAVESCAETHSGQRTGT	300	
Db	917	FLTERKLTVDGVVKLLRLRYNNEECRCNCTPGDPDKLYPFIYHAVESCAETHSGQRTGT	976	

## RESULT 2

US-09-511-477-4  
; Sequence 4, Application US/09511477  
; Patent No. 6337202  
; GENERAL INFORMATION:  
; APPLICANT: JACOBSON, Myron K.  
; APPLICANT: JACOBSON, Elaine L.  
; APPLICANT: AME, Jean-Christophe  
; APPLICANT: LIN, Winston  
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE (PARG) EN  
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV  
; TITLE OF INVENTION: THEREWITH  
; FILE REFERENCE: NIAD 201  
; CURRENT APPLICATION NUMBER: US/09/511,477  
; CURRENT FILING DATE: 2000-02-23  
; PRIOR APPLICATION NUMBER: 09/302,812  
; PRIOR FILING DATE: 1999-04-30  
; NUMBER OF SEQ ID NOS: 38  
; SEQ ID NO 4  
; LENGTH: 976  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
US-09-511-477-4

Query Match 100.0%; Score 1597; DB 4; Length 976;

Best Local Similarity 100.0%; Pred. No. 2.1e-171;  
Matches 300; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	KTLFCYFRVRVTEKPKTGLVTFTRQSLDFPWEKCEKPLTRLHVTVYEGTIENGQMLQV	60
DB	677	KTLFCYFRVRVTEKPKTGLVTFTRQSLDFPWEKCEKPLTRLHVTVYEGTIENGQMLQV	736
QY	61	DFANRFVGGVTSAGLVQVEIRFLINPELIIISRLFTFVLHDNECLIIITGEQYSEYTGVA	120
DB	737	DFANRFVGGVTSAGLVQVEIRFLINPELIIISRLFTFVLHDNECLIIITGEQYSEYTGVA	796
QY	121	ETYRWSRSHEDGSDRDCERRCTEIVADALHFRYLDQFVPEKMRRELKAYCGFLRPG	180
DB	797	ETYRWSRSHEDGSDRDCERRCTEIVADALHFRYLDQFVPEKMRRELKAYCGFLRPG	856
QY	181	VSENL SAVATGNWCGGAFGGDARLKALIQILAAAAERDVVYFTFGDSELMRDYSMHI	240
DB	857	VSENL SAVATGNWCGGAFGGDARLKALIQILAAAAERDVVYFTFGDSELMRDYSMHI	916
QY	241	FLTERKLTGVDVYKLLRYNNEECNCSTPGPDIKLYPFIYHAVESCAETHSGQRTGT	300
DB	917	FLTERKLTGVDVYKLLRYNNEECNCSTPGPDIKLYPFIYHAVESCAETHSGQRTGT	976

## RESULT 3

US-09-511-507-4  
; Sequence 4, Application US/09511507  
; Patent No. 6395543  
; GENERAL INFORMATION:  
; APPLICANT: JACOBSON, Myron K.  
; APPLICANT: JACOBSON, Elaine L.  
; APPLICANT: AME, Jean-Christophe  
; APPLICANT: LIN, Winston  
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE (PARG) EN  
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV  
; TITLE OF INVENTION: THEREWITH  
; FILE REFERENCE: NIAD 201  
; CURRENT APPLICATION NUMBER: US/09/511,507  
; CURRENT FILING DATE: 2000-02-23  
; PRIOR APPLICATION NUMBER: 09/302,812  
; PRIOR FILING DATE: 1999-04-30  
; NUMBER OF SEQ ID NOS: 38  
; SEQ ID NO 4  
; LENGTH: 976  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:

Query Match 93.7%; Score 1497; DB 4; Length 968;

Best Local Similarity 93.7%; Pred. No. 4.3e-160;  
Matches 281; Conservative 7; Mismatches 12; Indels 0; Gaps 0;

QY	1	KTLFCYFRVRVTEKPKTGLVTFTRQSLDFPWEKCEKPLTRLHVTVYEGTIENGQMLQV	60
DB	669	KTLFCYFRVRVTEKPKTGLVTFTRQSLDFPWEKCEKPLTRLHVTVYEGTIENGQMLQV	728
QY	61	DFANRFVGGVTSAGLVQVEIRFLINPELIIISRLFTFVLHDNECLIIITGEQYSEYTGVA	120
DB	729	DFANRFVGGVTSAGLVQVEIRFLINPELIIISRLFTFVLHDNECLIIITGEQYSEYTGVA	788
QY	121	ETYRWSRSHEDGSDRDCERRCTEIVADALHFRYLDQFVPEKMRRELKAYCGFLRPG	180
DB	789	ETYRWSRSHEDGSEKDWQRRCCTEIVADALHFRYLDQFVPEKMRRELKAYCGFLRPG	848
QY	181	VSENL SAVATGNWCGGAFGGDARLKALIQILAAAAERDVVYFTFGDSELMRDYSMHI	240
DB	849	VSENL SAVATGNWCGGAFGGDARLKALIQILAAAAERDVVYFTFGDSELMRDYSMHT	908
QY	241	FLTERKLTGVDVYKLLRYNNEECNCSTPGPDIKLYPFIYHAVESCAETHSGQRTGT	300

## US-09-511-507-4

Query Match 100.0%; Score 1597; DB 4; Length 976;  
Best Local Similarity 100.0%; Pred. No. 2.1e-171;  
Matches 300; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	KTLFCYFRVRVTEKPKTGLVTFTRQSLDFPWEKCEKPLTRLHVTVYEGTIENGQMLQV	60
DB	677	KTLFCYFRVRVTEKPKTGLVTFTRQSLDFPWEKCEKPLTRLHVTVYEGTIENGQMLQV	736
QY	61	DFANRFVGGVTSAGLVQVEIRFLINPELIIISRLFTFVLHDNECLIIITGEQYSEYTGVA	120
DB	737	DFANRFVGGVTSAGLVQVEIRFLINPELIIISRLFTFVLHDNECLIIITGEQYSEYTGVA	796
QY	121	ETYRWSRSHEDGSDRDCERRCTEIVADALHFRYLDQFVPEKMRRELKAYCGFLRPG	180
DB	797	ETYRWSRSHEDGSDRDCERRCTEIVADALHFRYLDQFVPEKMRRELKAYCGFLRPG	856
QY	181	VSENL SAVATGNWCGGAFGGDARLKALIQILAAAAERDVVYFTFGDSELMRDYSMHI	240
DB	857	VSENL SAVATGNWCGGAFGGDARLKALIQILAAAAERDVVYFTFGDSELMRDYSMHI	916
QY	241	FLTERKLTGVDVYKLLRYNNEECNCSTPGPDIKLYPFIYHAVESCAETHSGQRTGT	300
DB	917	FLTERKLTGVDVYKLLRYNNEECNCSTPGPDIKLYPFIYHAVESCAETHSGQRTGT	976

## RESULT 4

US-09-302-812-6  
; Sequence 6, Application US/09302812B  
; Patent No. 6333148  
; GENERAL INFORMATION:  
; APPLICANT: JACOBSON, Myron K.  
; APPLICANT: JACOBSON, Elaine L.  
; APPLICANT: AME, Jean-Christophe  
; APPLICANT: LIN, Winston  
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE (PARG) EN  
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV  
; TITLE OF INVENTION: THEREWITH  
; FILE REFERENCE: NIAD 201  
; CURRENT APPLICATION NUMBER: US/09/302,812B  
; CURRENT FILING DATE: 1999-04-30  
; EARLIER APPLICATION NUMBER: 60/083,768  
; EARLIER FILING DATE: 1998-05-01  
; NUMBER OF SEQ ID NOS: 38  
; SEQ ID NO 6  
; LENGTH: 968  
; TYPE: PRT  
; ORGANISM: Mus musculus  
; FEATURE:

## US-09-302-812-6

Query Match 93.7%; Score 1497; DB 4; Length 968;

Best Local Similarity 93.7%; Pred. No. 4.3e-160;  
Matches 281; Conservative 7; Mismatches 12; Indels 0; Gaps 0;

QY	1	KTLFCYFRVRVTEKPKTGLVTFTRQSLDFPWEKCEKPLTRLHVTVYEGTIENGQMLQV	60
DB	669	KTLFCYFRVRVTEKPKTGLVTFTRQSLDFPWEKCEKPLTRLHVTVYEGTIENGQMLQV	728
QY	61	DFANRFVGGVTSAGLVQVEIRFLINPELIIISRLFTFVLHDNECLIIITGEQYSEYTGVA	120
DB	729	DFANRFVGGVTSAGLVQVEIRFLINPELIIISRLFTFVLHDNECLIIITGEQYSEYTGVA	788
QY	121	ETYRWSRSHEDGSDRDCERRCTEIVADALHFRYLDQFVPEKMRRELKAYCGFLRPG	180
DB	789	ETYRWSRSHEDGSEKDWQRRCCTEIVADALHFRYLDQFVPEKMRRELKAYCGFLRPG	848
QY	181	VSENL SAVATGNWCGGAFGGDARLKALIQILAAAAERDVVYFTFGDSELMRDYSMHI	240
DB	849	VSENL SAVATGNWCGGAFGGDARLKALIQILAAAAERDVVYFTFGDSELMRDYSMHT	908
QY	241	FLTERKLTGVDVYKLLRYNNEECNCSTPGPDIKLYPFIYHAVESCAETHSGQRTGT	300



Db	858	VSSENLSAVATGNWCGAFGGDARLKALTIQLAAVAERDVVYTFGDSFLMRDIYSMT	917
Qy	241	FLTERKLTVDGVYKLLRYNEECNCSTPGDPIKLYPIFYHAYESCAETADHSGQRT	299
Db	918	FLTERKLTGEVYKLLRYNEECNCSTPGDPIKLYPIFYHAYESTOTTNOPGQRT	976

RESULT 8  
 US-09-511-477-2  
 ; Sequence 2, Application US/09511477  
 ; Patent No. 6337202  
 ; GENERAL INFORMATION:  
 ; APPLICANT: JACOBSON, Myron K.  
 ; APPLICANT: JACOBSON, Elaine L.  
 ; APPLICANT: AME, Jean-Christophe  
 ; APPLICANT: LIN, Winston  
 ; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE (PARG) EN-  
 ; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV  
 ; TITLE OF INVENTION: THEREWITH  
 ; FILE REFERENCE: NIAID 201  
 ; CURRENT APPLICATION NUMBER: US/09/511,477  
 ; CURRENT FILING DATE: 2000-02-23  
 ; PRIOR APPLICATION NUMBER: 09/302,812  
 ; PRIOR FILING DATE: 1999-04-30  
 ; NUMBER OF SEQ ID NOS: 38  
 ; SEQ ID NO 2  
 ; LENGTH: 977  
 ; TYPE: PRT  
 ; ORGANISM: Bos taurus  
 ; FEATURE:  
 ; US-09-511-477-2

Query Match	93.5%	Score 1493;	DB 4;	Length 977;
Best Local Similarity	94.0%;	Pred. No. 1.2e-159;		
Matches 281;	Conservative 7;	Mismatches 11;	Indels 0;	Gaps 0;
Qy	1	KTLCFYRRRVTEKKPTGLVTFRTQSLDDPPWERCCKPLTRLHVYTGTTIEENGQGLMQV	60	
Db	678	KTLCFYRRRVTEKKPTGLVTFRTQSLDDPPWERCCKLLTFLHVYTGTTIEENGQGLMQV	737	
Qy	61	DFANRFVGGGVTSAGLVQBEIRFLINPELLISRLFTFVLVDHNECLITGTQEYSEYTGYA	120	
Db	738	DFANRFVGGGVTSAGLVQBEIRFLINPELLISRLFTFVLVDHNECLITGTQEYSEYTGYA	797	
Qy	121	ETYNWSRSHEDGSRDDCERCCTEIVADLHFRFYRLDQFVPEKMRRLNKAYCGFLRPG	180	
Db	798	ETYNWSRSHEDGSRDDQWRETTEIVADLHFRFYRLDQFVPEKIRRLNKAYCGFLRPG	857	
Qy	181	VSSNLISAVATGNWCGAFGGDARLKALIQILAAAAAERDVVYFTFGDSELMRDYIYSHHI	240	
Db	858	VSSNLISAVATGNWCGAFGGDARLKALIQILAAAVAEERDVVYFTFGDSELMRDYIYSHMT	917	
Qy	241	FLTERKLTJGDVYIKLLRLRYNEECRNGSTPGDPDKLFFPIYHAVESCAETADHSQRTG	299	
Db	918	FLTERKLTJGCEVYIKLLRYNEECRNGSTPGDPDKLFFPIYHAVESCTOTNDPGORTG	976	

RESULT 9  
 US-09-511-507-2  
 ; Sequence 2, Application US/09511507  
 ; Patent No. 6395543  
 ; GENERAL INFORMATION:  
 ; APPLICANT: JACOBSON, Myron K.  
 ; APPLICANT: JACOBSON, Elaine L.  
 ; APPLICANT: AME, Jean-Christophe  
 ; APPLICANT: LIN, Winston  
 ; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE (PARG) EN  
 ; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV  
 ; TITLE OF INVENTION: THEREWITH  
 ; FILE REFERENCE: NIAD 201  
 ; CURRENT APPLICATION NUMBER: US/09/511.507  
 ; CURRENT FILING DATE: 2000-02-23

```

; PRIOR APPLICATION NUMBER: 09/302,812
; PRIOR FILING DATE: 1999-04-30
; NUMBER OF SEQ ID NOS: 38
; SEQ ID NO 2
; LENGTH: 977
; TYPE: PRT
; ORGANISM: Bos taurus
; FEATURE:
;
US-09-511-507-2

          Query Match          93.5%; Score 1493; DB 4; Length 977;
          Best Local Similarity 94.0%; Pred. No. 1.2e-159;
          Matches 281; Conservative 7; Mismatches 11; Indels 0; Gaps 0;

Qy      1   KTLFCYFRRVTEKKPTGLVTFTTQSLDDFPWEKRCCKPLRLHLVHYTGTTIENGCGMLQV 60
Db      678 KTLFCYFRRVTEKKPTGLVTFTTQSLDDFPWEKRCCKLLRLHLVHYTGTTIENGCGMLQV 737

Qy      61   DFANRFVGGGVTSAGLVQVEIRFLINPELLIISRLFTVLDHNECLLIITGTEQSEYTGYA 120
Db      738 DFANRFVGGGVTSAGLVQVEIRFLINPELLIISRLFTVLDHNECLLIITGTEQSEYTGYA 797

Qy      121  ETRYWRSRSHEDGSRDDCERCCTEIVAI DALHFRRYLDQFVPEKMRRELKAYCGFTLRPG 180
Db      798 ETRYWARSHEDRGSRDDWQRRTTEIVAI DALHFRRYLDQFVPEKIRRELKAYCGFTLRPG 857

Qy      181  VSSENISAVATGNWGCAGFGDARLKALQILAAAAAARDVVYFTFGDSLMDRIYSMHI 240
Db      858 VSSENISAVATGNWGCAGFGDARLKALQILAAA VAERDVVYFTFGDSLMDRIYSMHT 917

Qy      241  FLTERKLTVDGVYKLLRLYYNECRNCSTGGPDIKLYPFYIHAVESCAETADHSQRTG 299
Db      918 FLTERKLTGVEYVKLLRLYYNECRNCSTGGPDIKLYPFYIHAVESCTOTTNPGQRTG 976

```

RESULT 10  
US-09-302-812-8  
; Sequence 8, Application US/09302812B  
; Patent No. 6331148  
; GENERAL INFORMATION:  
; APPLICANT: JACOBSON, Myron K.  
; APPLICANT: JACOBSON, Elaine L.  
; APPLICANT: AM, Jean-Christophe  
; APPLICANT: LIM, Winscon  
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE (PARG) ENVI  
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIVE  
; TITLE OF INVENTION: THEREWITH  
; FILE REFERENCE: NIA2 201  
; CURRENT APPLICATION NUMBER: US/09/302,812B  
; CURRENT FILING DATE: 1999-04-30  
; EARLIER APPLICATION NUMBER: 60/083,768  
; EARLIER FILING DATE: 1998-05-01  
; NUMBER OF SEQ ID NOS: 38  
; SEQ ID NO 8  
; LENGTH: 768  
; TYPE: PRT  
; ORGANISM: Drosophila melanogaster  
; FEATURE:  
; US-09-302-812-8

Query Match 38.8%; Score 619.5; DB 4; Length 768;  
Best Local Similarity 46.4%; Pred. No. 6.3e-61;  
Matches 140; Conservative 41; Mismatches 90; Indels 31; Gaps 8

Qy 1 KTLFCYFRV--TEKK---PTGLVTTTQS--LEDPEWERCEKPL--TRLHVTVEGTL 50  
Db 297 KCMHYFRVCPETERDASNVPTGVVTVRRSGUPEHLIDWSQAAPLGDVPLHVDAGEL 356  
Qy 51 EENGQMLQVDFANRFVGGVTSAGLVQVEIRFLINPELLIISRLFEVLIDHNECLLIITGT 110  
Db 357 EDEGLGLQVDFANKYLGGLVGHGCVQVEIRFVFCIELLVGKLFTECLRPFEALVWLGA 416  
Qy 111 EQVSYTGYAETIRWSRSHEDSGSRDCCRRCTEIVADALHFRRYLDQVPEKKRRELN 170

Db 417 ERYSNYTGAGSFGWNGFNEDSTPRDSSGRQTAIWAIDALHFAQSHQYREDLMERELN 476  
QY 171 KAYCGFLR-----PGVSSENLSAVATGNWCGAFGGDARLKALIQILAAAAAERDVVVF 224  
Db 477 KAYIGFVHMVTPPPG-----VATGNWCGAFGGDSYLLKALLQMLVCAQLGRPLAYY 528  
QY 225 TFGDSELMRDIYSMHIFLTERKLTVDGVYKLLRLRYNEECRNCSTPGP-----DIKLYP 278  
Db 529 TFGNVEFRDDFHEMMLFRNDGTTVQQLWS-ILRSYSLRIKESKKEPRENKASKKLYD 587  
QY 279 FI 280  
Db 588 FI 589

RESULT 11  
US-09-511-477-8  
; Sequence 8, Application US/09511477  
; Patent No. 6337202  
; GENERAL INFORMATION:  
; APPLICANT: JACOBSON, Myron K.  
; APPLICANT: JACOBSON, Elaine L.  
; APPLICANT: AME, Jean-Christophe  
; APPLICANT: LIN, Winston  
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE (PARG) ENZ  
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV  
; FILE REFERENCE: NIAD 201  
; CURRENT APPLICATION NUMBER: US/09/511.477  
; CURRENT FILING DATE: 2000-02-23  
; PRIOR APPLICATION NUMBER: 09/302,812  
; PRIOR FILING DATE: 1999-04-30  
; NUMBER OF SEQ ID NOS: 38  
; SEQ ID NO 8  
; LENGTH: 768  
; TYPE: PRT  
; ORGANISM: Drosophila melanogaster  
; FEATURE:  
US-09-511-477-8

Query Match 38.8%; Score 619.5; DB 4; Length 768;  
Best Local Similarity 46.4%; Pred. No. 6.3e-61;  
Matches 140; Conservative 41; Mismatches 90; Indels 31; Gaps 8;  
QY 1 KTLFCYFRRV--TEKK-----PTGLVTFTRQS--LEDPEWERCEKPL--TRIHHVTEGTI 50  
Db 297 KCMHYFRRCVPTERDASNVPTGVVTFVRRSGLPEHLIDWSQSAAPLGVDVPLHVDAGTI 356  
QY 51 EENGQGMLOVDFAFRFVGGVTSAGLVQVEIRFLINPELIIISRLFTVLDHNECLIITGT 110  
Db 357 EDEGIGLLQVDFANKYLGGLGVGHGCVQVEIRFVCPGLLVGKLFTECLRPFEALVMLGA 416  
QY 111 EQYSEYTGAYETRWKSHEDSGDRCRRTCTEIVAIDALHFRFRLDQVFPKMRRELN 170  
Db 417 ERYSNYTGAGSFGWNGFNEDSTPRDSSGRQTAIWAIDALHFAQSHQYREDLMERELN 476  
QY 171 KAYCGFLR-----PGVSSENLSAVATGNWCGAFGGDARLKALIQILAAAAAERDVVVF 224  
Db 477 KAYIGFVHMVTPPPG-----VATGNWCGAFGGDSYLLKALLQMLVCAQLGRPLAYY 528  
QY 225 TFGDSELMRDIYSMHIFLTERKLTVDGVYKLLRLRYNEECRNCSTPGP-----DIKLYP 278  
Db 529 TFGNVEFRDDFHEMMLFRNDGTTVQQLWS-ILRSYSLRIKESKKEPRENKASKKLYD 587  
QY 279 FI 280  
Db 588 FI 589

RESULT 12  
US-09-511-507-8  
; Sequence 8, Application US/09511507

; Patent No. 6395543  
; GENERAL INFORMATION:  
; APPLICANT: JACOBSON, Myron K.  
; APPLICANT: JACOBSON, Elaine L.  
; APPLICANT: AME, Jean-Christophe  
; APPLICANT: LIN, Winston  
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE (PARG) ENZ  
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV  
; FILE REFERENCE: NIAD 201  
; CURRENT APPLICATION NUMBER: US/09/511.507  
; CURRENT FILING DATE: 2000-02-23  
; PRIOR APPLICATION NUMBER: 09/302,812  
; PRIOR FILING DATE: 1999-04-30  
; NUMBER OF SEQ ID NOS: 38  
; SEQ ID NO 8  
; LENGTH: 768  
; TYPE: PRT  
; ORGANISM: Drosophila melanogaster  
; FEATURE:  
US-09-511-507-8

Query Match 38.8%; Score 619.5; DB 4; Length 768;  
Best Local Similarity 46.4%; Pred. No. 6.3e-61;  
Matches 140; Conservative 41; Mismatches 90; Indels 31; Gaps 8;  
QY 1 KTLFCYFRRV--TEKK-----PTGLVTFTRQS--LEDPEWERCEKPL--TRIHHVTEGTI 50  
Db 297 KCMHYFRRCVPTERDASNVPTGVVTFVRRSGLPEHLIDWSQSAAPLGVDVPLHVDAGTI 356  
QY 51 EENGQGMLOVDFAFRFVGGVTSAGLVQVEIRFLINPELIIISRLFTVLDHNECLIITGT 110  
Db 357 EDEGIGLLQVDFANKYLGGLGVGHGCVQVEIRFVCPGLLVGKLFTECLRPFEALVMLGA 416  
QY 111 EQYSEYTGAYETRWKSHEDSGDRCRRTCTEIVAIDALHFRFRLDQVFPKMRRELN 170  
Db 417 ERYSNYTGAGSFGWNGFNEDSTPRDSSGRQTAIWAIDALHFAQSHQYREDLMERELN 476  
QY 171 KAYCGFLR-----PGVSSENLSAVATGNWCGAFGGDARLKALIQILAAAAAERDVVVF 224  
Db 477 KAYIGFVHMVTPPPG-----VATGNWCGAFGGDSYLLKALLQMLVCAQLGRPLAYY 528  
QY 225 TFGDSELMRDIYSMHIFLTERKLTVDGVYKLLRLRYNEECRNCSTPGP-----DIKLYP 278  
Db 529 TFGNVEFRDDFHEMMLFRNDGTTVQQLWS-ILRSYSLRIKESKKEPRENKASKKLYD 587  
QY 279 FI 280  
Db 588 FI 589

RESULT 13  
US-09-302-812-10  
; Sequence 10, Application US/09302812B  
; Patent No. 6333148  
; GENERAL INFORMATION:  
; APPLICANT: JACOBSON, Myron K.  
; APPLICANT: JACOBSON, Elaine L.  
; APPLICANT: AM, Jean-Christophe  
; APPLICANT: LIN, Winston  
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE (PARG) ENZ  
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV  
; FILE REFERENCE: NIAD 201  
; CURRENT APPLICATION NUMBER: US/09/302,812B  
; CURRENT FILING DATE: 1999-04-30  
; EARLIER APPLICATION NUMBER: 60/083,768  
; EARLIER FILING DATE: 1998-05-01  
; NUMBER OF SEQ ID NOS: 38  
; SEQ ID NO 10  
; LENGTH: 726  
; TYPE: PRT  
; ORGANISM: Caenorhabditis elegans

```
FEATURE:
US-09-302-812-10
Query Match      23.3%; Score 371.5; DB 4; Length 726;
Best Local Similarity 35.3%; Pred. No. 6.8e-33;
Matches 94; Conservative 49; Mismatches 104; Indels 19; Gaps 8;

QY 1 KTLFCYFRVRVTEKKTGLVTF--TROSLEDF-PWEKCEKPLTRLHVT--YEGTIEHQ 55
Db 456 KFLFTYFDKMSDPPDGVAFSLTKMDKDTFNEWK--DKLRLSLPEVEFFDEMLIEDTA 513
QY 56 GMLQVDPANRFVGGVTSAGLVQEEIRFLNPELIIISRLFTVLDHNECLIIITGTEQYSE 115
Db 514 LCTQVDPANEHLOGGVNLHSGVQEEIRFLMCPMMVGMMLCEKMKQLEAISIVGAVVFS 573
QY 116 YTGAYETVRW-----SRSHEDGSE-RDDCERRCTEIVADALHFR-----YLDQFVPEK 164
Db 574 YTGYGHTLKWAELOPNHRSQNTNEFRDRLRVEITAIIDALFKGSKLDCQTEQLNKAN 633
QY 165 MRRELKAYCGFLRPGVSSSENLSAVATGNWCGGAFGGDARLKALIQILAAAAERDVVYF 224
Db 634 IIREMKKASIGFMSQGPFTNI-PIVTGWGCGAFNGDKPLKPLFIQIVIAAGVADRLPHFC 692
QY 225 TFGDSELMRDIYSMHIFLTERKLTVG 250
Db 693 SFGEPELAACKKIIERMKQKDVTLG 718

RESULT 14
US-09-511-477-10
; Sequence 10, Application US/09511477
; Patent No. 6337202
; GENERAL INFORMATION:
; APPLICANT: JACOBSON, Myron K.
; APPLICANT: JACOBSON, Elaine L.
; APPLICANT: AME, Jean-Christophe
; APPLICANT: LIN, Winston
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE (PARG) EN
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV
; FILE REFERENCE: NIAD 201
; CURRENT APPLICATION NUMBER: US/09/511,477
; CURRENT FILING DATE: 2000-02-23
; PRIOR APPLICATION NUMBER: 09/302,812
; PRIOR FILING DATE: 1999-04-30
; NUMBER OF SEQ ID NOS: 38
; SEQ ID NO 10
; LENGTH: 726
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
; FEATURE:

US-09-511-477-10
Query Match      23.3%; Score 371.5; DB 4; Length 726;
Best Local Similarity 35.3%; Pred. No. 6.8e-33;
Matches 94; Conservative 49; Mismatches 104; Indels 19; Gaps 8;

QY 1 KTLFCYFRVRVTEKKTGLVTF--TROSLEDF-PWEKCEKPLTRLHVT--YEGTIEHQ 55
Db 456 KFLFTYFDKMSDPPDGVAFSLTKMDKDTFNEWK--DKLRLSLPEVEFFDEMLIEDTA 513
QY 56 GMLQVDPANRFVGGVTSAGLVQEEIRFLNPELIIISRLFTVLDHNECLIIITGTEQYSE 115
Db 514 LCTQVDPANEHLOGGVNLHSGVQEEIRFLMCPMMVGMMLCEKMKQLEAISIVGAVVFS 573
QY 116 YTGAYETVRW-----SRSHEDGSE-RDDCERRCTEIVADALHFR-----YLDQFVPEK 164
Db 574 YTGYGHTLKWAELOPNHRSQNTNEFRDRLRVEITAIIDALFKGSKLDCQTEQLNKAN 633
QY 165 MRRELKAYCGFLRPGVSSSENLSAVATGNWCGGAFGGDARLKALIQILAAAAERDVVYF 224
Db 634 IIREMKKASIGFMSQGPFTNI-PIVTGWGCGAFNGDKPLKPLFIQIVIAAGVADRLPHFC 692
QY 225 TFGDSELMRDIYSMHIFLTERKLTVG 250
Db 693 SFGEPELAACKKIIERMKQKDVTLG 718

RESULT 15
US-09-511-507-10
; Sequence 10, Application US/09511507
; Patent No. 6395543
; GENERAL INFORMATION:
; APPLICANT: JACOBSON, Myron K.
; APPLICANT: JACOBSON, Elaine L.
; APPLICANT: AME, Jean-Christophe
; APPLICANT: LIN, Winston
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE (PARG) EN
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV
; FILE REFERENCE: NIAD 201
; CURRENT APPLICATION NUMBER: US/09/511,507
; CURRENT FILING DATE: 2000-02-23
; PRIOR APPLICATION NUMBER: 09/302,812
; PRIOR FILING DATE: 1999-04-30
; NUMBER OF SEQ ID NOS: 38
; SEQ ID NO 10
; LENGTH: 726
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
; FEATURE:

US-09-511-507-10
Query Match      23.3%; Score 371.5; DB 4; Length 726;
Best Local Similarity 35.3%; Pred. No. 6.8e-33;
Matches 94; Conservative 49; Mismatches 104; Indels 19; Gaps 8;

QY 1 KTLFCYFRVRVTEKKTGLVTF--TROSLEDF-PWEKCEKPLTRLHVT--YEGTIEHQ 55
Db 456 KFLFTYFDKMSDPPDGVAFSLTKMDKDTFNEWK--DKLRLSLPEVEFFDEMLIEDTA 513
QY 56 GMLQVDPANRFVGGVTSAGLVQEEIRFLNPELIIISRLFTVLDHNECLIIITGTEQYSE 115
Db 514 LCTQVDPANEHLOGGVNLHSGVQEEIRFLMCPMMVGMMLCEKMKQLEAISIVGAVVFS 573
QY 116 YTGAYETVRW-----SRSHEDGSE-RDDCERRCTEIVADALHFR-----YLDQFVPEK 164
Db 574 YTGYGHTLKWAELOPNHRSQNTNEFRDRLRVEITAIIDALFKGSKLDCQTEQLNKAN 633
QY 165 MRRELKAYCGFLRPGVSSSENLSAVATGNWCGGAFGGDARLKALIQILAAAAERDVVYF 224
Db 634 IIREMKKASIGFMSQGPFTNI-PIVTGWGCGAFNGDKPLKPLFIQIVIAAGVADRLPHFC 692
QY 225 TFGDSELMRDIYSMHIFLTERKLTVG 250
Db 693 SFGEPELAACKKIIERMKQKDVTLG 718
```

Search completed: May 26, 2004, 18:49:26  
Job time: 10.6131 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 26, 2004, 18:46:21 ; Search time 25.0302 Seconds

(without alignments)

3343.018 Million cell updates/sec

Title: US-09-302-812-4\_COPY\_677\_977

Perfect score: 1597

Sequence: 1 KTLFCYFRRVTEKKPTGLVT.....YHAVESCAETADHSQRTGT 300

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1149313 seqs, 278921704 residues

Total number of hits satisfying chosen parameters: 1149313

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:\*

- 1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*
- 2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pep.\*
- 3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep.\*
- 4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep.\*
- 5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep.\*
- 6: /cgn2\_6/ptodata/1/pubpaa/PCTUS\_PUBCOMB.pep.\*
- 7: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep.\*
- 8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep.\*
- 9: /cgn2\_6/ptodata/1/pubpaa/US09A\_PUBCOMB.pep.\*
- 10: /cgn2\_6/ptodata/1/pubpaa/US09B\_PUBCOMB.pep.\*
- 11: /cgn2\_6/ptodata/1/pubpaa/US09C\_PUBCOMB.pep.\*
- 12: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep.\*
- 13: /cgn2\_6/ptodata/1/pubpaa/US10A\_PUBCOMB.pep.\*
- 14: /cgn2\_6/ptodata/1/pubpaa/US10B\_PUBCOMB.pep.\*
- 15: /cgn2\_6/ptodata/1/pubpaa/US10C\_PUBCOMB.pep.\*
- 16: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep.\*
- 17: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep.\*
- 18: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1597	100.0	976	9	US-09-973-451-4
2	1497	93.7	968	9	US-09-973-451-6
3	1493	93.5	977	9	US-09-973-451-2
4	619.5	38.8	768	9	US-09-973-451-8
5	489	30.6	546	12	US-10-425-114-60000
6	371.5	23.3	726	9	US-09-973-451-10
7	191.5	12.0	200	12	US-10-424-599-184988
8	188	11.8	180	12	US-10-424-599-156445
9	167	10.5	31	9	US-09-973-451-11
10	149	9.3	29	9	US-09-973-451-12
11	99.5	6.2	300	12	US-10-424-599-233915
12	97.5	6.1	395	12	US-10-282-122A-51256
13	94.5	5.9	730	15	US-10-369-493-15591
14	92.5	5.8	7349	14	US-10-314-657-46
15	91.5	5.7	261	14	US-10-247-671-134

16	91.5	5.7	261	15	US-10-435-696-27	Sequence 27, Appl
17	89	5.6	205	9	US-09-879-957-16	Sequence 16, Appl
18	88	5.5	442	15	US-10-369-493-11707	Sequence 11707, A
19	85	5.3	581	9	US-09-887-586A-48	Sequence 48, Appl
20	85	5.3	581	9	US-09-903-012-48	Sequence 48, Appl
21	85	5.3	581	10	US-09-900-797-48	Sequence 48, Appl
22	85	5.3	581	12	US-09-893-820-48	Sequence 48, Appl
23	85	5.3	581	13	US-10-041-007-18	Sequence 18, Appl
24	84.5	5.3	982	14	US-10-156-761-11410	Sequence 11410, A
25	83.5	5.2	153	12	US-10-282-122A-56462	Sequence 56462, A
26	83.5	5.2	505	9	US-09-604-287A-478	Sequence 478, App
27	83.5	5.2	505	9	US-09-604-287A-485	Sequence 485, App
28	83.5	5.2	505	10	US-09-551-621-478	Sequence 478, App
29	83.5	5.2	505	13	US-10-007-805-478	Sequence 478, App
30	83.5	5.2	505	13	US-10-007-805-485	Sequence 485, App
31	83.5	5.2	505	14	US-10-076-622-478	Sequence 478, App
32	83.5	5.2	505	14	US-10-076-622-485	Sequence 485, App
33	83.5	5.2	505	14	US-10-124-805-478	Sequence 478, App
34	83.5	5.2	505	14	US-10-124-805-485	Sequence 485, App
35	82.5	5.2	294	12	US-10-424-599-249573	Sequence 249573, A
36	82.5	5.2	452	12	US-10-425-114-41682	Sequence 41682, A
37	82.5	5.2	624	12	US-10-282-122A-51658	Sequence 51658, A
38	82.5	5.2	854	14	US-10-244-488-13	Sequence 13, Appl
39	82.5	5.2	996	15	US-10-369-493-18505	Sequence 18505, A
40	81.5	5.1	275	12	US-10-425-114-69783	Sequence 69783, A
41	81	5.1	329	15	US-10-369-493-233	Sequence 233, App
42	81	5.1	645	12	US-10-282-122A-53367	Sequence 53367, A
43	81	5.1	905	13	US-10-114-893-127	Sequence 127, App
44	81	5.1	984	15	US-10-369-493-20751	Sequence 20751, A
45	80	5.0	275	12	US-10-282-122A-52681	Sequence 52681, A

## ALIGNMENTS

RESULT 1  
US-09-973-451-4  
; Sequence 4, Application US/09973451  
; Patent No. US2002013228A1  
; GENERAL INFORMATION:  
; APPLICANT: JACOBSON, Myron K.  
; APPLICANT: JACOBSON, Elaine L.  
; APPLICANT: AM, Jean-Christophe  
; APPLICANT: LIN, Wianston  
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE  
; TITLE OF INVENTION: (PARG) ENZYMES,  
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV.  
; TITLE OF INVENTION: THEREWITH  
; FILE REFERENCE: NIAD 201  
; CURRENT APPLICATION NUMBER: US/09/973,451  
; PRIOR FILING DATE: 2001-10-09  
; PRIOR APPLICATION NUMBER: US/09/302,812  
; PRIOR FILING DATE: 1999-04-30  
; PRIOR APPLICATION NUMBER: 60/083,768  
; PRIOR FILING DATE: 1998-05-01  
; NUMBER OF SEQ ID NOS: 38  
; SEQ ID NO 4  
; LENGTH: 976  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
US-09-973-451-4  
Query Match 100.0%; Score 1597; DB 9; Length 976;  
Best Local Similarity 100.0%; Pred. No. 1.3e-171;  
Matches 300; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 KTLFCYFRRVTEKKPTGLVTFTTQSLDEPWECEKPLTRHLVTEGTEENGCQMLQV 60  
Db 677 KTLFCYFRRVTEKKPTGLVTFTTQSLDEPWECEKPLTRHLVTEGTEENGCQMLQV 736  
QY 61 DFANRFGGVTSGALVQEEIRFLINPELIIIRLFTFVLDHNECLITGTQYSEYTGVA 120  
|||||

Db 737 DFANFVGGVTSAGLVQBEIRFLINPELIIISRLFTFVLDHNECLIITGTEQSYBTGYA 796  
QY 121 EYRWSRSHEDSGERDDCRRCTEIVADALHFRYLDQFVPEKMRRELINLKAYCGFLRPG 180  
Db 797 EYRWSRSHEDSGERDDCRRCTEIVADALHFRYLDQFVPEKMRRELINLKAYCGFLRPG 856  
QY 181 VSENLAVATGNWCGAFGGDARLKALIQIILAAAAERDVVYFTFGDSELMRDYISMHI 240  
Db 857 VSENLAVATGNWCGAFGGDARLKALIQIILAAAAERDVVYFTFGDSELMRDYISMHI 916  
QY 241 FLTERKLTVDGVYKLLRYNNEECNCSPPGDIKLYPIYHVESCAETADHSGORTG 300  
Db 917 FLTERKLTVDGVYKLLRYNNEECNCSPPGDIKLYPIYHVESCAETADHSGORTG 976

## RESULT 2

US-09-973-451-6

; Sequence 6, Application US/09973451  
; Patent No. US20020132328A1

; GENERAL INFORMATION:

; APPLICANT: JACOBSON, Myron K.

; APPLICANT: JACOBSON, Elaine L.

; APPLICANT: AM, Jean-Christophe

; APPLICANT: LIN, Winston

; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE

; TITLE OF INVENTION: (PARG) ENZYMES,

; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV

; TITLE OF INVENTION: THEREWITH

; FILE REFERENCE: NIAD 201

; CURRENT APPLICATION NUMBER: US/09/973,451

; PRIOR FILING DATE: 2001-10-09

; PRIOR APPLICATION NUMBER: US/09/302,812

; PRIOR FILING DATE: 1999-04-30

; PRIOR APPLICATION NUMBER: 60/083,768

; PRIOR FILING DATE: 1998-05-01

; NUMBER OF SEQ ID NOS: 38

; SEQ ID NO 6

; LENGTH: 968

; TYPE: PRT

; ORGANISM: Mus musculus

; FEATURE:

US-09-973-451-6

Query Match

Best Local Similarity 93.7%; Score 1497; DB 9; Length 968;

Matches 281; Conservative 7; Mismatches 12; Indels 0; Gaps 0;

QY 1 KTLFCYFRVTEKKTGLVTFTRQSLDPPWERCEKPLTRLHVYEGTIEENGOMLOV 60  
Db 669 KTLFCYFRVTEKKTGLVTFTRQSLDPPWERCEKPLTRLHVYEGTIEENGOMLOV 728  
QY 61 DFANFVGGVTSAGLVQBEIRFLINPELIIISRLFTFVLDHNECLIITGTEQSYBTGYA 120  
Db 729 DFANFVGGVTSAGLVQBEIRFLINPELIIISRLFTFVLDHNECLIITGTEQSYBTGYA 788  
QY 121 EYRWSRSHEDSGERDDCRRCTEIVADALHFRYLDQFVPEKMRRELINLKAYCGFLRPG 180  
Db 789 EYRWSRSHEDSGERDDCRRCTEIVADALHFRYLDQFVPEKMRRELINLKAYCGFLRPG 848  
QY 181 VSENLAVATGNWCGAFGGDARLKALIQIILAAAAERDVVYFTFGDSELMRDYISMHI 240  
Db 849 VSENLAVATGNWCGAFGGDARLKALIQIILAAAAERDVVYFTFGDSELMRDYISMHI 908  
QY 241 FLTERKLTVDGVYKLLRYNNEECNCSPPGDIKLYPIYHVESCAETADHSGORTG 300  
Db 909 FLTERKLTVDGVYKLLRYNNEECNCSPPGDIKLYPIYHVESCAETADHSGORTG 968

## RESULT 3

US-09-973-451-2

; Sequence 2, Application US/09973451

; Patent No. US20020132328A1

; GENERAL INFORMATION:

; APPLICANT: JACOBSON, Myron K.  
; APPLICANT: JACOBSON, Elaine L.  
; APPLICANT: AM, Jean-Christophe  
; APPLICANT: LIN, Winston  
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE  
; TITLE OF INVENTION: (PARG) ENZYMES,  
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV  
; TITLE OF INVENTION: THEREWITH  
; FILE REFERENCE: NIAD 201  
; CURRENT APPLICATION NUMBER: US/09/973,451  
; CURRENT FILING DATE: 2001-10-09  
; PRIOR APPLICATION NUMBER: US/09/302,812  
; PRIOR FILING DATE: 1999-04-30  
; PRIOR APPLICATION NUMBER: 60/083,768  
; PRIOR FILING DATE: 1998-05-01  
; NUMBER OF SEQ ID NOS: 38  
; SEQ ID NO 2  
; LENGTH: 977  
; TYPE: PRT  
; ORGANISM: Bos taurus  
; FEATURE:

US-09-973-451-2

Query Match

Best Local Similarity 93.5%; Score 1493; DB 9; Length 977;

Matches 281; Conservative 7; Mismatches 11; Indels 0; Gaps 0;

QY 1 KTLFCYFRVTEKKTGLVTFTRQSLDPPWERCEKPLTRLHVYEGTIEENGOMLOV 60  
Db 678 KTLFCYFRVTEKKTGLVTFTRQSLDPPWERCEKPLTRLHVYEGTIEENGOMLOV 737  
QY 61 DFANFVGGVTSAGLVQBEIRFLINPELIIISRLFTFVLDHNECLIITGTEQSYBTGYA 120  
Db 738 DFANFVGGVTSAGLVQBEIRFLINPELIIISRLFTFVLDHNECLIITGTEQSYBTGYA 797  
QY 121 EYRWSRSHEDSGERDDCRRCTEIVADALHFRYLDQFVPEKMRRELINLKAYCGFLRPG 180  
Db 798 EYRWSRSHEDSGERDDCRRCTEIVADALHFRYLDQFVPEKMRRELINLKAYCGFLRPG 857  
QY 181 VSENLAVATGNWCGAFGGDARLKALIQIILAAAAERDVVYFTFGDSELMRDYISMHI 240  
Db 858 VSENLAVATGNWCGAFGGDARLKALIQIILAAAAERDVVYFTFGDSELMRDYISMHI 917  
QY 241 FLTERKLTVDGVYKLLRYNNEECNCSPPGDIKLYPIYHVESCAETADHSGORTG 299  
Db 918 FLTERKLTVDGVYKLLRYNNEECNCSPPGDIKLYPIYHVESCAETADHSGORTG 976

## RESULT 4

US-09-973-451-8

; Sequence 8, Application US/09973451

; Patent No. US20020132328A1

; GENERAL INFORMATION:

; APPLICANT: JACOBSON, Myron K.

; APPLICANT: JACOBSON, Elaine L.

; APPLICANT: AM, Jean-Christophe

; APPLICANT: LIN, Winston

; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE

; TITLE OF INVENTION: (PARG) ENZYMES,

; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV

; TITLE OF INVENTION: THEREWITH

; FILE REFERENCE: NIAD 201

; CURRENT APPLICATION NUMBER: US/09/973,451

; CURRENT FILING DATE: 2001-10-09

; PRIOR APPLICATION NUMBER: US/09/302,812

; PRIOR FILING DATE: 1999-04-30

; PRIOR APPLICATION NUMBER: 60/083,768

; PRIOR FILING DATE: 1998-05-01

; NUMBER OF SEQ ID NOS: 38

; SEQ ID NO 8

; LENGTH: 768

; TYPE: PRT

; ORGANISM: Drosophila melanogaster

; FEATURE:

US-09-973-451-8

Query Match 38.8%; Score 619.5; DB 9; Length 768;

Best Local Similarity 46.4%; Pred. No. 1e-60; Indels 31; Gaps 8;

Matches 140; Conservative 41; Mismatches 90; Indels 31; Gaps 8;

QY 1 KTLFCYFRRV--TEKK-----PTGLVTFTRQS--LEDFPEWERCEKPL--TRLHVTYEGTI 50

DB 297 KCIHMYFRRVCPTERDASNVFTGVVTFVRRSGLPEHLIDWSQSAAPLGDVPLHVAEGTI 356

QY 51 BENGQMLQVDFANRFVGGVTSAGLVQEEIRFLINPELLIISRLFTFVLDHNECLIIITG 110

DB 357 EDEIGLQVDFANKYLGGLGSHGCVQEEIRFVICPELLVGLKFTCELRPFALVMLGA 416

QY 111 EYSEYTGVAETVRSRSHEDGSDRDCERCTEIVADALHFRYLDQFVPEKMRRELN 170

DB 417 ERYSNYTGAGSFWGNGFSTFRDSSGRQTAIVADALHFAQSHQYREDLMRELN 476

QY 171 KAYCGFLR-----PGVSSSENLSAVATGNWCGGAFGGDARLKALIQILAAAAARDVVVF 224

DB 477 KAYIGFVHMVTPPG-----VATGNWCGGAFGGDSYLLKALLQLMCAQLGRPLAYY 528

QY 225 TFGDSELMRDYISHMIFLTERKLTVDGVYKLLRYNEECNCSCTPGP-----DIKLYP 278

DB 529 TFGNVEFRDDFHEWMLLFRNDGTTVQQLWS--ILRSYSLRIKEKSKPEPRENKASKKLYD 587

QY 279 FI 280

DB 588 FI 589

RESULT 5

US-10-425-114-60000

; Sequence 60000, Application US/10425114

; Publication No. US20040034888A1

; GENERAL INFORMATION:

; APPLICANT: Liu, Jingdong

; APPLICANT: Zhou, Yihua

; APPLICANT: Kovalic, David K.

; APPLICANT: Screen, Steven E

; APPLICANT: Tabaska, Jack E

; APPLICANT: Cao, Yongwei

; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With

; FILE REFERENCE: 38-21(5313)B

; CURRENT APPLICATION NUMBER: US/10/425,114

; CURRENT FILING DATE: 2003-04-28

; NUMBER OF SEQ ID NOS: 73128

; SEQ ID NO 60000

; LENGTH: 546

; TYPE: PRT

; ORGANISM: Zea mays

; FEATURE:

; OTHER INFORMATION: Clone ID: LIB3689-262-E12\_FLI.pep

US-10-425-114-60000

Query Match

Best Local Similarity 38.4%; Score 489; DB 12; Length 546;

Matches 113; Conservative 46; Mismatches 101; Indels 34; Gaps 5;

QY 1 KTLFCYFRRVTEKKPTGLVTFTRQS-----EDFPEWERCEKPLTRLHVTYEGT 49

DB 226 RCLVHYFERVTDNPTGLSVFERKVLPRRALSDGVPPYDIIHAWVASSAPLCQTFVSSGF 285

QY 50 IRENGQMLQVDFANRFVGGVTSAGLVQEEIRFLINPELLIISRLFTFVLDHNECLIIITG 109

DB 286 IEDDEQALQVDFANKYLGGLGSHGCVQEEIRFVICPELLVGLKFTCELRPFALVMLGA 345

QY 110 EYSEYTGVAETVRSRSHEDGSDRDCERCTEIVADALHFRYLDQFVPEKMRREL 169

DB 346 AERFSQWYGGSSRFVGDYLDTKPFDMSGRRRTIIVADALDPCARL-HYESGCLLREV 404

QY 170 NKAYCGFLR-----PCVSSSENLSAVATGNWCGGAFGGDARLKALI 209

DB 405 NKATCGFDQSKHLYAKLFQDLHNKDDFSSINSSEYVGSTGNWCGGAFGNPEIKSMI 464

QY 210 QILAAAAARDVV-YFTFGDSELMRDYISHMIFLTERKLTVDGVYKLLRYNE 262

DB 465 QWIAAQAALRPVNYTFEDVSLQR-LEEVIQWIRLHGTVGLWHMLMEYSQ 517

RESULT 6

US-09-973-451-10

; Sequence 10, Application US/09973451

; Patent No. US20020132328A1

; GENERAL INFORMATION:

; APPLICANT: JACOBSON, Myron K.

; APPLICANT: JACOBSON, Elaine L.

; APPLICANT: AM, Jean-Christophe

; APPLICANT: LIN, Winston

; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE

; TITLE OF INVENTION: (PARG) ENZYMES

; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV

; TITLE OF INVENTION: THEREWITH

; FILE REFERENCE: NIAID 201

; CURRENT APPLICATION NUMBER: US/09/973,451

; CURRENT FILING DATE: 2001-10-09

; PRIOR APPLICATION NUMBER: US/09/302,812

; PRIOR FILING DATE: 1999-04-30

; PRIOR APPLICATION NUMBER: 60/083,768

; PRIOR FILING DATE: 1998-05-01

; NUMBER OF SEQ ID NOS: 38

; SEQ ID NO 10

; LENGTH: 726

; TYPE: PRT

; ORGANISM: Caenorhabditis elegans

; FEATURE:

US-09-973-451-10

Query Match

Best Local Similarity 23.3%; Score 371.5; DB 9; Length 726;

Matches 94; Conservative 49; Mismatches 104; Indels 19; Gaps 8;

QY 1 KTLFCYFRRVTEKKPTGLVTF--TRQSLDF--PEWERCERKPLTRLHVT--YEGTIENGG 55

DB 456 KFLTFYFDKMSDPPDGAVSFRUTKMDKDTFNEWK--DKKLRSLEPVEFFDEMLIEDTA 513

QY 56 GMLQVDFANRFVGGVTSAGLVQEEIRFLINPELLIISRLFTFVLDHNECLIIITG 115

DB 514 LCTQVDFANRHLGGVNLHSGVQEEIRFLMCPENMVMGLLCEKMKQLEAISIVGAYVFS 573

QY 116 YTGVAETVRYW-----SRSHEDGSE--RDDCERRCTEIVADALHFR-----YLDQFVPEK 164

DB 574 YTGCGHTLKWAELOPNHSRQNTNEFDRFGLRVERIADAILFKGSKLDCQTEQLNKAN 633

QY 165 MRBELNAYCGFLRPGVSSSENLSAVATGNWCGGAFGGDARLKALIQILAAAAARDVVVF 224

DB 634 IIREMKKASIGFWSQGFKFTNI--PIVTGWWCGGAFNGDKPLKFTIIQVIAAGVADRPLHFC 692

QY 225 TFGDSELMRDYISHMIFLTERKLTVG 250

DB 693 SFGEPELAACKCKIIERMKQKQVTLG 718

RESULT 7

US-10-424-599-184988

; Sequence 184988, Application US/10424599

; Publication No. US20040031072A1

; GENERAL INFORMATION:

; APPLICANT: La Rosa Thomas J

; APPLICANT: Kovalic David K

; APPLICANT: Zhou Yihua

; APPLICANT: Cao Yongwei

; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With

; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement

```
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 184988
; LENGTH: 200
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(200)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_138059C.1.pep
US-10-424-599-184988

Query Match      12.0%; Score 191.5; DB 12; Length 200;
Best Local Similarity 36.0%; Pred. No. 5.6e-13;
Matches 40; Conservative 19; Mismatches 41; Indels 11; Gaps 2;

QY      6 YFRRVTEKPTGLVTFRTQSLR-----DFPE---MERCEKPLRLHVTYEGTIEENG 54
Db      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
90 YGQRISSEPKGIVSFERKVLFPKDSIHISYEDANFMTSAIPLCRFEVHSGSLIEDQS 149
Db      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
55 QQMLQVDFANRFVGGVTSAGLVQBEIRFLINPELIISRLFTFVLDHNECL 105
Db      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
150 SGAEEVDLANKYLGGAALRGCGVQBEIRFPWSPLEAGMLFLPAMADNEAI 200

RESULT 8
US-10-424-599-156445
; Sequence 156445, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated with
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 156445
; LENGTH: 180
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_112291C.1.pep
US-10-424-599-156445

Query Match      11.8%; Score 188; DB 12; Length 180;
Best Local Similarity 30.1%; Pred. No. 1.2e-12;
Matches 49; Conservative 23; Mismatches 41; Indels 50; Gaps 4;

QY      168 ELNKAYCGFL-----RP-----179
Db      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
7 EINKAFCGFLYCKYQPKILQENGCTSAIPYATSTSMETDEGEISNHNKITSQNDYH 66

QY      180 GVSSENLSAVATGNWCGAFGGDARLKALIQILAAAAARD-VVYFTFGDSSELMRDYISM 238
Db      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
67 GMDQGNIGVATGNWCGAFGDPVKTIQWLAASQALRPPIAYTTG-LEALQSLDEV 125

QY      239 HIFLTERKLTGVDVVKLLRYNECRNCSTPGPDIKLYPFIY 281
Db      ::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
126 AHWILSORWTVGDLWNMLIEYSINRSKGETNVGFLQWLLPSIY 168

RESULT 9
US-09-973-451-11
; Sequence 11, Application US/09973451
; Patent No. US20020132328A1
```

```
; GENERAL INFORMATION:
; APPLICANT: JACOBSON, Myron K.
; APPLICANT: JACOBSON, Elaine L.
; APPLICANT: AM, Jean-Christophe
; APPLICANT: LIN, Winston
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE
; TITLE OF INVENTION: (PARG) ENZYMES
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV
; TITLE OF INVENTION: THEREWITH
; FILE REFERENCE: NIAD 201
; CURRENT APPLICATION NUMBER: US/09/973,451
; CURRENT FILING DATE: 2001-10-09
; PRIOR APPLICATION NUMBER: US/09/302,812
; PRIOR FILING DATE: 1999-04-30
; PRIOR APPLICATION NUMBER: 60/083,768
; PRIOR FILING DATE: 1998-05-01
; NUMBER OF SEQ ID NOS: 38
; SEQ ID NO 11
; LENGTH: 31
; TYPE: PRT
; ORGANISM: Bos taurus
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_112291C.1.pep
US-09-973-451-11

Query Match      10.5%; Score 167; DB 9; Length 31;
Best Local Similarity 100.0%; Pred. No. 2.1e-11;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      94 LFTFVLDHNECLIIITGTQYSEYTGVAETFR 124
Db      |||||
1 LFTFVLDHNECLIIITGTQYSEYTGVAETFR 31

RESULT 10
US-09-973-451-12
; Sequence 12, Application US/09973451
; Patent No. US20020132328A1
; GENERAL INFORMATION:
; APPLICANT: JACOBSON, Myron K.
; APPLICANT: JACOBSON, Elaine L.
; APPLICANT: AM, Jean-Christophe
; APPLICANT: LIN, Winston
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE
; TITLE OF INVENTION: (PARG) ENZYMES
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV
; TITLE OF INVENTION: THEREWITH
; FILE REFERENCE: NIAD 201
; CURRENT APPLICATION NUMBER: US/09/973,451
; CURRENT FILING DATE: 2001-10-09
; PRIOR APPLICATION NUMBER: US/09/302,812
; PRIOR FILING DATE: 1999-04-30
; PRIOR APPLICATION NUMBER: 60/083,768
; PRIOR FILING DATE: 1998-05-01
; NUMBER OF SEQ ID NOS: 38
; SEQ ID NO 12
; LENGTH: 29
; TYPE: PRT
; ORGANISM: Bos taurus
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_112291C.1.pep
US-09-973-451-12

Query Match      9.3%; Score 149; DB 9; Length 29;
Best Local Similarity 96.6%; Pred. No. 2.1e-09;
Matches 28; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      172 AYCGLRPGVSSSENLSAVATGNWCGAFG 200
Db      |||||
1 AYCGLRPGVSSSENLSAVATGNWCGAFG 29

RESULT 11
US-10-424-599-233915
; Sequence 233915, Application US/10424599
```



```

Db 316 EYRGVFEKXHALSRFPQKIDYSEPDIGETVBLRLKLSRYEKHNKVKYTEVALTAARLS 375
Qy 153 FRYLDQVPEPM-----RRELKAYCGFLR---PGVSEN 185
Db 376 ARFINDRLHPDKAIDVIDEAGAAQRILPKSRQKIIGRQIEQVIAGIARIIPPONVSSDD 435
Qy 186 LSAVATGNWCGGA--FGDARLKALIQILAAAAE-----RDVVYFTF-----GDSLM 232
Db 436 RNKULTDRDLKALVFGQDAIDALTSIAKMARSGLNTCKPIGSLFGTGVGKTEVA 495
Qy 233 RDI-YSMHI 240
Db 496 RQLAYTLGI 504

RESULT 14
US-10-314-657-46
; Sequence 46, Application US/10314657
; Publication No. US2003017588A1
; GENERAL INFORMATION:
; APPLICANT: SHEN, Ben
; APPLICANT: CHENG, Yi-Qiang
; APPLICANT: TANG, Gong-Li
; TITLE OF INVENTION: Discrete Acyltransferases Associated with Type I Polyketide
; TITLE OF INVENTION: Syntheses and Methods of Use
; FILE REFERENCE: 054030-0021
; CURRENT APPLICATION NUMBER: US/10/314,657
; CURRENT FILING DATE: 2002-12-09
; PRIOR APPLICATION NUMBER: PCT/US02/08937
; PRIOR FILING DATE: 2002-03-22
; PRIOR APPLICATION NUMBER: US 60/278,935
; PRIOR FILING DATE: 2001-03-26
; NUMBER OF SEQ ID NOS: 214
; SOFTWARE: Patent in version 3.2
; SEQ ID NO 46
; LENGTH: 7349
; TYPE: PRT
; ORGANISM: Streptomyces atroolivaceus
US-10-314-657-46

Query Match 5.8%; Score 92.5; DB 14; Length 7349;
Best Local Similarity 22.5%; Pred. No. 21;
Matches 62; Conservative 31; Mismatches 89; Indels 93; Gaps 12;

Qy 40 TRLHYTE--GTHENGQMLQVDFANRFVGGVTSAGLVQEEI-----RFLI 85
Db 641 TRLVTHETLAVPSGNGAGVLREDGVYLVTTGGGSLAALLVRLVTRGPVRLVLTGRSAP 700
Qy 86 NPELI-----ISRLFTEVLDHNECL--IITGTEQYSEYTGVAETYRWSRSHED 131
Db 701 GPFLTQRIEGWRRGAEVTHVRGDVAHTDDVLAAYTCARETYGRIDG---VFHCAGSVDD 757
Qy 132 GS-ERDDCERRCT-----EIVADILHP-----RR 155
Db 758 GNFRRKDPERSAAVLAALKVAGTRNLDEATADDGLAFFALFSSVSASVANPGQADYAYGNA 817
Qy 156 YLDQVPEKWRRELKAYCGFLRPGVS-----SENLSAVATGNWCGGAFG 200
Db 818 FMEHFAEQRAARA-----DRPGVSVAVGWLPLWADGMVRVSEDVLRSAADTSLGLHALP 869
Qy 201 GDARLKALIQILAAAAARDVVYFTFGDSLMRDI 235
Db 870 ADAGLDALFGLI-SGNAPRAVV--TYGDQRIAEI 901

```

## RESULT 15

```

US-10-247-671-134
; Sequence 134, Application US/10247671
; Publication No. US20030194721A1
; GENERAL INFORMATION:
; APPLICANT: Mikita, Thomas
; APPLICANT: Shiffman, Dov
; APPLICANT: Porter, Gordon, J.

```

```

; APPLICANT: Kaser, Matthew R.
; TITLE OF INVENTION: GENES EXPRESSED IN TREATED FOAM CELLS
; FILE REFERENCE: PR-0050 US
; CURRENT APPLICATION NUMBER: US/10/247,671
; CURRENT FILING DATE: 2002-09-18
; PRIOR APPLICATION NUMBER: 60/323,784
; PRIOR FILING DATE: 2001-09-19
; NUMBER OF SEQ ID NOS: 186
; SOFTWARE: PERL Program
; SEQ ID NO 134
; LENGTH: 261
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. US20030194721A1 2616727CD1
US-10-247-671-134

```

```

Query Match 5.7%; Score 91.5; DB 14; Length 261;
Best Local Similarity 22.6%; Pred. No. 0.19;
Matches 53; Conservative 32; Mismatches 80; Indels 69; Gaps 9;

Qy 12 EKKPTGLVTFTRQS---LEDPEWERC-EKELTRLHVTYEGTIEEN-CQGLQVDFANRF 66
Db 48 EKPYCNAHYPKQSFVTWVADTPENLRKQOSELSQVRYKEEFKNGKKG-----97
Qy 67 VGGGVTSAGLVQEEIRFLINPELIIISRLFTVLDHNECLITGTEQYSEYTGVAETYRWS 126
Db 98 -----FSVADTPELQRIKKTDQIISNIKYHEFEKS 129
Qy 127 RSHEDGSRDDCERRCTEIVADILHFFRYLDQFVP-----EKKRELKAYCG 175
Db 130 RMGPSGGEGMEPERDSQ---DGSSYRRLPELQCPHPIPTSPVYQOPQOPVAQSYGG 185
Qy 176 FLRPGVSENLSAVATGNWCGGAFGADARLKALIQILAAAAARDVVYFTFGDS 229
Db 186 YKEP-AAPVSIQRSAPG-----GGGKRYRAVDY---SAADEDEVSFQDGD 228

```

Search completed: May 26, 2004, 19:18:56  
Job time : 26.0302 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model  
Run on: May 26, 2004, 18:23:35 ; Search time 8.25272 Seconds  
(without alignments)  
3496.721 Million cell updates/sec

Title: US-09-302-812-4\_COPY\_677\_977  
Perfect score: 1597  
Sequence: 1 KTLFCVFRVTEKPTGLVT.....YHAVESCAETADHSQRTGT 300

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues  
Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries  
Database : PIR 78: \*  
1: pir1: \*  
2: pir2: \*  
3: pir3: \*  
4: pir4: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	406	25.4	997	B84726	probable poly(ADP-ribose) glycohydrolase [imported] - Arabidopsis thaliana
2	371.5	23.3	726	T21138	hypothetical prote
3	254.5	15.9	364	A84726	probable poly(ADP-ribose) glycohydrolase [imported] - Arabidopsis thaliana
4	91.5	5.7	261	S68234	Laap-1 protein - h
5	91.5	5.7	491	S05408	keratin, type II,
6	90	5.6	1194	S70415	DNA-directed RNA p
7	88.5	5.5	2135	T14602	variant-specific s
8	88	5.5	448	AB2740	acetyl-CoA carboxy
9	88	5.5	448	H97520	biotin carboxylase
10	87.5	5.5	391	AB2322	hypothetical prote
11	86.5	5.4	153	G90831	probable endopepti
12	86	5.4	153	APBPML	endopeptidase (EC
13	86	5.4	153	APB21	endopeptidase (EC
14	85	5.3	523	S03572	DNA-directed RNA p
15	85	5.3	523	B84416	DNA-directed RNA p
16	84.5	5.3	317	F82672	ATP sulfurylase, s
17	84.5	5.3	397	T35713	probable oxidoredu
18	83.5	5.2	153	B64788	hypothetical prote
19	83.5	5.2	3036	T18995	hypothetical prote
20	83	5.2	729	A69202	conserved hypotet
21	82.5	5.2	624	D97307	molecular chaperon
22	82.5	5.2	681	S27868	glvr-1 protein - m
23	82.5	5.2	996	D86872	beta-galactosidase
24	82	5.1	256	S16255	hypothetical prote
25	82	5.1	282	A97214	uncharacterized co
26	81.5	5.1	600	A45112	major parafagella
27	81	5.1	602	AH2583	ATP-dependent DNA
28	81	5.1	602	H97365	DNA helicase XF138
29	80.5	5.0	477	C69318	reductase, iron-su

hypothetical prote  
indolepyruvate oxi  
ABC transporter-li  
herbicide safener  
hypothetical prote  
competence-induced  
phosphohexosemutas  
translation elonga  
trehalose-6-phosph  
insulysin (EC 3.4.  
hypothetical prote  
serine/threonine-s  
beta-galactosidase  
probable protein k  
hypothetical prote  
hypothetical prote

ALIGNMENTS

RESULT 1

B84726  
probable poly(ADP-ribose) glycohydrolase [imported] - Arabidopsis thaliana  
C;Species: Arabidopsis thaliana (mouse-ear cress)  
C;Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 02-Feb-2001  
C;Accession: B84726  
R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;  
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.  
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J  
Nature 402, 761-768, 1999  
A;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.  
A;Reference number: A84420; MUID:20083487; PMID:10617197  
A;Accession: B84726  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-997 <STO>  
A;Cross-references: GB:AE002093; NID:g4887749; PIDN:AAD32285.1; GSPDB:GN00139  
C;Genetics:  
A;Gene: At2g31870  
A;Map position: 2

Query Match 25.4%; Score 406; DB 2; Length 997;  
Best Local Similarity 34.5%; Pred. No. 2.9e-29;  
Matches 105; Conservative 33; Mismatches 72; Indels 94; Gaps 9;  
QY 6 YFRVTEKPTGLVTFTRQ--SLEDPEWERCEKPLTRLHVTEGTEIENGQGLQVDFEA 63  
DB 196 YPERFCSCVPIGIVSFERKITAAPDAFWSKSDVSL-----YQ-----PDNALEVDFA 243  
QY 64 NRFVGGVTSAGLVQBEIRFLINPELIISRLFTFVLDHNECLIIITGEQVSEYTGVAETV 123  
DB 244 NKVLGGSLSRGCVQBEIRFMINPELLAGMLFLPMDNDNEATEIVGAERFSCYGVASSF 303  
QY 124 RMRSHEDSGSRDDCERRCTEIVADAL-----HPRVLDQFVPEKMRRELKAVCGFL 177  
DB 304 RFAGEYIDKKAMPFKRRTRIVADALCTPKMRHFKDIC-----LLREINKALCGFL 356  
QY 178 R----- 178  
DB 357 NCSKAWEHQNI PMDEGNDIEQLVRNGRDSGLLRTETTASHRTPLNDVNRKPPANNLR 416  
QY 179 -----PGVSSSENI--SAVATENMGCGAFGGDARKALKALIIQLAAAAARDVV- YTFGDSGL 231  
DB 417 DFVVEGVDNEDHDDGVAIVGNWCGVFGDPKELKATIQWLAAASQTRRPISYITFG-VEA 475  
QY 232 MRDI 235  
DB 476 LRNL 479

RESULT 2

T21138

hypothetical protein F20C5.1 - *Caenorhabditis elegans*  
 C:Species: *Caenorhabditis elegans*  
 C>Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 04-Mar-2000  
 C:Accession: T21138  
 R:Matthews, P.  
 submitted to the EMBL Data Library, December 1995  
 A:Reference number: Z19381  
 A:Accession: T21138  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-726 <WIL>  
 A:Cross-references: EMBL:Z68161; PIDN:CAA92299.1; GSPDB:GN00022; CESP:F20C5.1  
 A:Experimental source: clone F20C5  
 C:Genetics:  
 A:Gene: CESP.F20C5.1  
 A:Map position: 4  
 A:Introns: 15/3; 162/3; 219/3; 294/1; 316/3; 347/3; 403/3; 442/2; 536/3; 578/1; 675/1  
 C:Superfamily: *Caenorhabditis elegans* hypothetical protein F20C5.1

Query Match 23.3%; Score 371.5; DB 2; Length 726;  
 Best Local Similarity 35.3%; Pred. No. 3e-26;  
 Matches 94; Conservative 49; Mismatches 104; Indels 19; Gaps 8;

QY 1 KTLFCYFRVTKKPTGLVTF--TROSLEDF-PEWERCERKPLRLHVT--YEGTIEENQ 55  
 Db 456 KFLFYFDKSMDDPPDGVAFRLTKMDKDTFNEWK--DKLRLSPEVEFFDEMLIEDTA 513  
 QY 56 GMLQVDFANRFGVGGVTSAGLVOEIRFLINPELIIISRLFTVLDHNECLITGTQVSE 115  
 Db 514 LCTQVDFANEHLGGVILNHSQVEIRFLMCPPEVMVGMLLCEKMKQLEAISVGVAFVSS 573  
 QY 116 YTGAYAEYRW-----SRSHEDGSE-RDDCERRCTEIVADLHFR-----YLDQFVPEK 164  
 Db 574 YTGCHTLKWAELQPNHSQNTNEFRGRRLRVETIADILFKGSKLDCQTEQLNKAN 633  
 QY 165 MERELNKAYCGFLRPGVSSSENISAVATGNWCGAGFGDARLKALIQILAAAAAERDVVVF 224  
 Db 634 IIREMKASIGFMSQGPKEFTNI-PIVTGWGCGAFNGDKPLAFIIQVIAAGVADRLPHEC 692  
 QY 225 IFGDSLMRDIYSMHIFLTERKLTGVG 250  
 Db 693 SFGEPELAAKCKKITERMKQKQVTLG 718

RESULT 3  
 A84726  
 probable poly (ADP-ribose) glycohydrolase [imported] - *Arabidopsis thaliana*  
 C:Species: *Arabidopsis thaliana* (mouse-ear cress)  
 C>Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 02-Feb-2001  
 C:Accession: A84726  
 R:Lin, X.; Kaul, S.; Rounslev, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;  
 M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.;  
 euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.  
 Nature 402, 761-768, 1999  
 A:Title: Sequence and analysis of chromosome 2 of the plant *Arabidopsis thaliana*.  
 A:Reference number: A84420; MUID:20083487; PMID:10617197  
 A:Accession: A84726  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-364 <STO>  
 A:Cross-references: GB:AE002093; NID:g4887750; PIDN:AAD32286.1; GSPDB:GN00139  
 C:Genetics:  
 A:Gene: At2g31860  
 A:Map position: 2

Query Match 15.9%; Score 254.5; DB 2; Length 364;  
 Best Local Similarity 28.3%; Pred. No. 9.5e-16;  
 Matches 76; Conservative 36; Mismatches 78; Indels 79; Gaps 7;

QY 1 KTLFCYFRVTKKPTGLVTFTRQ--SLR-----DFPWEKCEKPLRLHVTYGTTEE 52  
 Db 112 KCIHYFORLSSISPGFFVSFERKILSLRQDSSTLDEGFWGKSTVNLCPVEVRTSGLIED 171

QY 53 NGQGMLOVDENRFGVGGVTSAGLVOEIRFLINPELIIISRLFTVLDHNECLITGTQ 112  
 Db 172 QSVLEAVDFANKUGGALRGKCVQOEIRFMINPELVGMLFLTMEVTEIAEVVGAER 231  
 QY 113 YSEYTGAYAEYRWRSRSHEDGSDCERRC-TEIIVADLHFRYLDQFVPEKMRRLNK 171  
 Db 232 FSLYTG-----CFKAKTRIVADALR----- 253  
 QY 172 AYCGLRPGVSSSENISAVATGNWCGAGFGDARLKALIQILAAAAAERDVVVF 225  
 Db 254 -----HFGVS-----QTKLESLSVLILSSSGRPIRLYMGVSLQ 288  
 QY 226 -FGDSLMRDIYSMHIFLTERKLTGVGVV 253  
 Db 289 GIGDVVLAVELSSSLFENGURFHSNLY 317

RESULT 4  
 S68234  
 Lasp-1 protein - human  
 N:Alternate names: LIM and SH3 protein  
 C:Species: *Homo sapiens* (man)  
 C>Date: 15-Feb-1997 #sequence\_revision 13-Mar-1997 #text\_change 16-Jul-1999  
 C:Accession: S68234; S72556  
 R:Tomasetto, C.; Moog-Lutz, C.; Regnier, C.H.; Schreiber, V.; Basset, P.; Rio, M.C.  
 FEBS Lett. 373, 245-249, 1995  
 A:Title: Lasp-1 (MUN 50) defines a new LIM protein subfamily characterized by the assoc.  
 A:Reference number: S68234; MUID:96033982; PMID:7589475  
 A:Accession: S68234  
 A:Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-261 <TOM>  
 A:Cross-references: EMBL:X82456  
 R:Tomasetto, C.  
 submitted to the EMBL Data Library, November 1994  
 A:Reference number: S72556  
 A:Accession: S72556  
 A:Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-84, 'C', 86-90, 'G', 92-219, 'D', 221-261 <TO2>  
 A:Cross-references: EMBL:X82456; NID:g951273; PID:g951274  
 C:Superfamily: LIM metal-binding repeat homology; SH3 homology  
 F:5-56/Domain: LIM metal-binding repeat homology <LIM>  
 F:209-258/Domain: SH3 homology <SH3>

Query Match 5.7%; Score 91.5; DB 2; Length 261;  
 Best Local Similarity 22.6%; Pred. No. 0.92;  
 Matches 53; Conservative 32; Mismatches 80; Indels 69; Gaps 9;

QY 12 EKKPTGLVTFTRQS---LEDPEWERC-EKPLRLHVTVEGTIERN-GQGMLOVDENR 66  
 Db 48 EKKPCNNAHYKQSFMTVADTPENLRKQSELOSQVYKEEFKNKGK----- 97  
 QY 67 VGGVTSAGLVOEIRFLINPELIIISRLFTVLDHNECLITGTQVSEYGYAITYAWS 126  
 Db 98 -----FSVADTPQLRIKKTQDQISNIRYHEFEKS 129  
 QY 127 RSHEDGSDCERRCTEIVADLHFRYLDQFV-----EKMRELNKAYCG 175  
 Db 130 RMGPGSGGEMPERDSQ-----DGSSYRRLPEQQPHIPTSPVYQPOQPOVAQSYGG 185  
 QY 176 FLRPGVSSSENISAVATGNWCGAGFGDARLKALIQILAAAAAERDVVVFTRGDS 229  
 Db 186 YKEP-AAPVSIQSRAPG-----GGGKRYRVDY---SAADEDEVSFQDGT 228

RESULT 5  
 S05408  
 keratin, type II, component 7c, cytoskeletal - sheep (tentative sequence)  
 C:Species: *Ovis orientalis aries*, *Ovis ammon aries* (domestic sheep)  
 C>Date: 30-Jun-1991 #sequence\_revision 30-Jun-1991 #text\_change 31-Mar-2000  
 C:Accession: S05408  
 R:Sparrow, L.G.; Robinson, C.P.; McMahon, D.T.W.; Rubira, M.R.



Biochem. J. 261, 1015-1022, 1989

A;Title: The amino acid sequence of component 7c, a type II intermediate-filament protein  
A;Reference number: S05408; MUID:90026244; PMID:2803231  
A;Accession: S05408  
A;Molecule type: protein  
A;Residues: 1-491 <SPA>  
C;Superfamily: cytoskeletal keratin  
C;Keywords: coiled coil

Query Match 5.7%; Score 91.5; DB 2; Length 491;  
Best Local Similarity 23.9%; Pred. No. 2.1;  
Matches 62; Conservative 32; Mismatches 96; Indels 69; Gaps 13;

QY 50 IEENGQMLQVDFANRFGVGGVTSAGLVQERFL---INPELIISRLFTVLDHNECL 105  
DB 228 LEANSEALIQEIDFLRR-----LYQEEIRVLQANISDTSVIVKMDNSRDLNMDCI 277

QY 106 IITGTEQYSEVT--GYAETVWRSHEDSGSERDDCERRCTEIVAIDALH---FRRYLDQF 160  
DB 278 VASIKQYQDIASRSRAESWTRS-----KCEIKATVIRHGETLRRYKEEI 325

QY 161 -----VPEKMRRELKAYC--GFLRPGVS--SENLSAVATGNWCGAFGGDARLKALIQIL 212  
DB 326 NELNRVIRQLTAENVAKQNSKLEAAVTQAEQQGEVALNDARCKLAGLEALQKAKQDM 385

QY 213 AAAAARDVVYTFGDSLMRDYSHMIFLTERKLTVDGVYKLLRYNNECRNCSTPGP 272  
DB 386 ACILKE-----YQVWMSKGLDIE---IATYRRLLEG-----BEORLCRGVG- 425

QY 273 DIKLYPFIVHAVESCAETA 291  
DB 426 -----AVNVCVSSS 434

#### RESULT 6

S70415  
DNA-directed RNA polymerase (EC 2.7.7.6) II second largest chain - Euplotes octocarinatus  
C;Species: Euplotes octocarinatus  
C;Date: 28-Oct-1996 #sequence\_revision 13-Mar-1997 #text\_change 20-Jun-2000  
C;Accession: S70415

R;Kaufmann, J.; Florian, V.; Klein, A.  
Nucleic Acids Res. 20, 5985-5989, 1992  
A;Title: TGA cysteine codons and intron sequences in conserved and nonconserved position  
A;Reference number: S70412; MUID:93036587; PMID:1461731  
A;Accession: S70415  
A;Molecule type: DNA

A;Residues: 1-1194 <KAU>  
A;Cross-references: EMBL:X66453; NID:g11071; PIDN:CMA47069.1; PID:g2654279  
C;Genetics:

A;Gene: RPB2  
A;Genetic code: SGC9  
A;Introns: 88/1

C;Superfamily: DNA-directed RNA polymerase 132K polypeptide  
C;Keywords: DNA binding; nucleotidyltransferase; nucleus; transcription

Query Match 5.6%; Score 90; DB 2; Length 1194;  
Best Local Similarity 24.3%; Pred. No. 9.6;  
Matches 43; Conservative 28; Mismatches 78; Indels 28; Gaps 8;

QY 52 ENGQGMQLQVDFANRFGVGGVTSAGLVQEEIRFLINPELIISRLFTVLDHNECLIIITGTE 111  
DB 332 EESTPFLTQEDALSYIGSRGTAGVGRER-RIRHAREILKRHPLPHVSTKDSK-----ES 385

QY 112 QVSEYTGAEYTWRSRSHEDSGSERDCE-----RCCTEIVAIDALHPRRVLDFVPEKMR- 166  
DB 386 QKSYFTGYM-CHRLNLAHLRIKRIKEDRDHYGKKRLDMAGSLGMLPNLFRFRVKEAGRY 444

QY 167 -----RELNKAYCGFLRPGVSSENLS-AVATGNWCGAFGGDARLKALIQIL 212  
DB 445 LKAQADKGSNLITKAF-----KKDIISDLGLKALATGNWGTNA-AGEVSKTGVSQVL 496

#### RESULT 7

#### TI4602

variant-specific surface protein - malaria parasite (Plasmodium falciparum) (fragments)  
C;Species: Plasmodium falciparum  
C;Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 09-Jun-2000  
C;Accession: TI4602

R;Voss, T.S.; Felger, I.; Weiss, N.; Beck, H.P.  
submitted to the EMBL Data Library, February 1998  
A;Description: Identification of a conserved 5' flanking region of Plasmodium falciparum  
A;Reference number: Z18158  
A;Accession: TI4602

A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1-2135 <VOS>

A;Cross-references: EMBL:AF050740; NID:g2944094; PID:g2944095; PIDN:AAC05220.1  
C;Genetics:  
A;Gene: varph17

Query Match 5.5%; Score 88.5; DB 2; Length 2135;  
Best Local Similarity 23.7%; Pred. No. 29;  
Matches 37; Conservative 24; Mismatches 58; Indels 37; Gaps 7;

QY 118 GYAET----YRWSRSHEDSGSERDDCERRCTEIVAIDALHPRRYLDQF-----VPEKM 165  
DB 59 GNIETCSLEVKYKYPHNGGSDVSDKYPCKEL-----SRKYVDRFSDKIGGCYNEM 111

QY 166 RELNKAYCGFLRPGVSSSENLSAVATGNWCGAFGGDARLKALIQILAAAAARDVV--Y 223  
DB 112 RRDGIGACAPYRRLHLCHNLETIETNNVE---SNNAKHLLVDVCMAAKVEGDSIKTY 167

QY 224 FT-----FGDSLMRDYSHMIFLTERKLTVDGVYK 254  
DB 168 YTGHHQHKYDDSQLC-----TVLARSFADIGDIVR 196

#### RESULT 8

AB2740  
acetyl-CoA carboxylase, biotin carboxylase [imported] - Agrobacterium tumefaciens (strain  
C;Species: Agrobacterium tumefaciens  
C;Date: 11-Jan-2002 #sequence\_revision 11-Jan-2002 #text\_change 18-Nov-2002  
C;Accession: AB2740

R;Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, L.  
erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutyavin, T.; Levy, R.; Li, M.; McCrell  
; Karp, P.; Romero, P.; Zhang, S.  
Science 294, 2317-2323, 2001

A;Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm, J.  
ster, E.W.

A;Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.  
A;Reference number: AB2577; MUID:21608550; PMID:11743193

A;Accession: AB2740  
A;Status: preliminary  
A;Molecule type: DNA

A;Residues: 1-448 <KUR>  
A;Cross-references: GB:AE008688; PIDN:AAL42336.1; PID:g17739740; GSPDB:GN00186  
A;Experimental source: strain C58 (Dupont)

C;Genetics:  
A;Gene: accC

A;Map position: circular chromosome  
C;Superfamily: biotin carboxylase; biotin carboxylase homology

Query Match 5.5%; Score 88; DB 2; Length 448;  
Best Local Similarity 27.4%; Pred. No. 4;  
Matches 48; Conservative 16; Mismatches 59; Indels 52; Gaps 9;

QY 33 ERCEKPLRLHVTYGGTIE---ENQO-----GMLQVDFANRFGVGGVTSAGLVQEEIR 82  
DB 256 EICADAMRKLKVRGAGTIEFLYENGFEFYIEMNTRLQVEHP---VTEATGMDLVQEQIR 312

QY 83 FLINPELIISRLFTVLDHNECLIIITGTEQY-----ITGTEQY-----SEYTG 118  
DB 313 VASGQGLSVTQADIIEFHGHAIECRINAEDPRTPVSPGTLTYFHTPGGLGVGVDSGAYQG 372

QY 119 Y-AETVWRSHEDSGSERDDCERRCTEIVAIDALHPRRYLDQFVPEKMRREL 169

Db 373 YKIPYDSMIGKLVHGRDRDECIRR-----LRRALDFVVDGIKTTL 416

RESULT 9  
H97520  
biotin carboxylase (a chain of acetyl-CoA carboxylase (acc) [imported] - Agrobacterium tumefaciens  
C:Species: Agrobacterium tumefaciens  
C:Date: 30-Sep-2001 #sequence\_revision 30-Sep-2001 #text\_change 18-Nov-2002  
C:Accession: H97520  
A: Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman, R.; Liu, F.; Wollam, C.; Allinger, M.; Doughy, D.; Scott, C.; Lappas, C.; Markelz, B.; Science 294, 2323-2328, 2001  
A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tumefaciens  
A:Reference number: A97359; PMID:21608551; PMID:11743194  
A:Status: preliminary  
A:Accession: H97520  
A:Molecule type: DNA  
A:Residues: 1-448 <KUR>  
A:Cross-references: GB:AE007869; PIDN:BAK87121.1; PID:gl15156385; GSPDB:GN00169  
C:Genetics:  
A:Gene: AGR\_C\_2451  
A:Map position: circular chromosome  
C:Superfamily: biotin carboxylase; biotin carboxylase homology

Query Match 5.5%; Score 88; DB 2; Length 448;  
Best Local Similarity 27.4%; Pred. No. 4;  
Matches 48; Conservative 16; Mismatches 59; Indels 52; Gaps 9;

QY 33 ERCEPLRLVHTYEGTIE---ENGQ-----GMLQVDFANRFVGGVTSAGLVQEEIR 82  
DB 256 EICADAMKLYRGAGTIEFLYENGFEFYIENMTRLOVEHP---VTEALTGMIDVQEQIR 312  
QY 83 FLINPELLISRLFTVLDHN-ECLI-----ITGTEQY-----SEYTG 118  
DB 313 VASGGGLSVTDQADIEFHGAIECRINAEDPTFVSPGLTFTFTPGGLGVGVDSGAYQG 372

QY 119 Y-AETVRSRSH---DGSERDCCRRCTEIVDAIDALHFRYLDQFVPEKMEREL 169  
DB 373 YKIPYDSMIGKLVHGRDRDECIRR-----LRRALDFVVDGIKTTL 416

RESULT 10  
AB2322  
hypothetical protein alr4129 [imported] - Nostoc sp. (strain PCC 7120)  
C:Species: Nostoc sp. PCC 7120  
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120  
C:Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 09-Dec-2002  
C:Accession: AB2322  
A: Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi, R.; Nakazaki, N.; Shimo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S. DNA Res. 8, 205-213, 2001  
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena PCC 7120  
A:Reference number: AB1807; PMID:21595285; PMID:11759840  
A:Accession: AB2322  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-391 <KUR>  
A:Cross-references: GB:BA000019; PIDN:BAF75828.1; PID:gl17133264; GSPDB:GN00179  
A:Experimental source: strain PCC 7120  
C:Genetics:  
A:Gene: alr4129  
C:Superfamily: hypothetical protein AF0390

Query Match 5.5%; Score 87.5; DB 2; Length 391;  
Best Local Similarity 20.5%; Pred. No. 3.7;  
Matches 50; Conservative 33; Mismatches 96; Indels 65; Gaps 10;

QY 36 EKPLRLVHTYEGTIE---ENGQGLQVDFANRFVGGV-----TSAGLVQEEIRLINPE 88  
DB 140 EYPQIHLHAFSPQEQVFTIARDVGLGVVISALQWAGVNSLPGTAELVDDDEVRLICPE 199  
QY 89 LIISRLFTVLD-----HNECLIITG-TEQYSEYTGVAETVRSRSHEDGSRDDCER 140

Db 200 KINTATWLEIIGTAHVGLHTTTSTILSCHITPEKIGHLEKL-----RSLQQTAINHKYPA 256

QY 141 RCTEIVDAIDALHFRYLDQFVPEKMERELN-----KAYCGFLRPG-----180  
DB 257 RITEFIVLP-----FVQGEAPKSLRRVRGDRQDPTLADALLGAVARIYLGWIPNHOPS 310

QY 181 -VSENLSAVATGNWC-----GAFGGD-----ARLKALIQILAAAAAER 219  
DB 311 WKGLAGATEALTWGCNDIGGTLMEHITTMAGAVGTCTMEVATLQNAIASIGRPYQOR 370

QY 220 DVVY 223  
DB 371 DTLY 374

RESULT 11  
G90831  
probable endopeptidase [imported] - Escherichia coli (strain O157:H7, substrain RIMD 05)  
C:Species: Escherichia coli  
C:Date: 18-Jul-2001 #sequence\_revision 18-Jul-2001 #text\_change 03-Aug-2001  
C:Accession: G90831  
A: Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G. gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H. DNA Res. 8, 11-22, 2001  
A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genomic reference number: A99629; PMID:21156231; PMID:11256796  
A:Accession: G90831  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-153 <HAY>  
A:Cross-references: GB:BA000007; PIDN:BA835046.1; PID:gl13361087; GSPDB:GN00154  
A:Experimental source: strain O157:H7, substrain RIMD 0509952  
C:Genetics:  
A:Gene: ECs1623  
C:Superfamily: phage PA2 endopeptidase

Query Match 5.4%; Score 86.5; DB 2; Length 153;  
Best Local Similarity 23.7%; Pred. No. 1.3;  
Matches 46; Conservative 24; Mismatches 59; Indels 65; Gaps 9;

QY 90 IISRLFTVLDHNECLIITGTEQYSEYTGVAETVRSRSHEDGSRD-----DCER 140  
DB 7 IISALLI-----CIIVCLSWAVNHYRDNATYKAQR---DKNARELKLANVAITDMQM 56

QY 141 RCTEIVDAIDALHFRYLDQFVPEK-----RELN-KAYCGFLRPGVSENLSAVAT 191  
DB 57 RQDVAAALDADKYTKELADAKAENDALHDVAAAGRRRLHIKAVCOSVREATTASGVNNAAS 116

QY 192 GNWGGAFGGDARLKALIQILAAAAAARDVVYFTFGDSELMRDIYSMHIFLTERKLTVGD 251  
DB 117 -----PRL-----ADTAERD--YFT-----LREERLITWQK 139

QY 252 VYKLLRYNBEER 265  
DB 140 QLEGTQKYNEQCR 153

RESULT 12  
AFBPML  
endopeptidase (EC 3.4.-.-) - phage lambda  
C:Species: phage lambda  
C:Date: 13-Jun-1983 #sequence\_revision 13-Jun-1983 #text\_change 18-Jun-1999  
C:Accession: A94614; C92891; JN0749; A00906  
A: Daniels, D. submitted to the Nucleic Acid Sequence Database, September 1982  
A:Reference number: A94614  
A:Accession: A94614  
A:Molecule type: DNA  
A:Residues: 1-153 <DAN>  
A: Sanger, F.; Coulson, A.R.; Hong, G.F.; Hill, D.F.; Petersen, G.B. J. Mol. Biol. 162, 729-773, 1982  
A:Title: Nucleotide sequence of bacteriophage lambda DNA.  
A:Reference number: A92891; PMID:83189071; PMID:622115

[illegible]

QY 138 -CE--RRCTEIV-----AIDALHPR-----RYLDQFVPEKMRRELNK----- 171  
Db 325 LCRMAEACFELALGRREADDDKHANKRLKVSGLMKDLFTALNKLARDVKYQLERANM 384  
QY 172 -----AYCGFLRPGVSSNLS-AVATGNWCGAFG 200  
Db 385 RNRELTNTVVRSDVLTRELEHPIATGNWVGGRSG 419

Search completed: May 26, 2004, 18:47:32  
Job time : 11.2527 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: May 26, 2004, 16:13:47 ; Search time 189.999 Seconds  
(without alignments)  
13517.524 Million cell updates/sec

Title: US-09-302-812-1

Perfect score: 7242

Sequence: 1 accggaaagtgaagagcc.....aaattttcatttaacaaaaa 4070

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 2034082

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ n2p.model -DEV=xlpl  
-Q=/cgn2\_1/USPTO\_spool\_p/6333148/runat\_26052004\_150053\_5676/app\_query.fasta\_1.12437  
-DB=SPTREMBL 25 -QFMT=fastan -SUFFIX=n2p.rspt -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOALIGN=200 -THR\_SCORE=pet -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=ptco -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=6333148@cgn\_1\_1\_640\_@runat\_26052004\_150053\_5676 -NCPU=6 -ICPU=3  
-NO\_MMAP -LARGQUERY -NEG\_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

SPTREMBL\_25.\*  
1: sp archaea.\*  
2: sp bacteria.\*  
3: sp fungi.\*  
4: sp human.\*  
5: sp invertebrate.\*  
6: sp mammal.\*  
7: sp mhc.\*  
8: sp organelle.\*  
9: sp phage.\*  
10: sp plant.\*  
11: sp rodent.\*  
12: sp virus.\*  
13: sp vertebrate.\*  
14: sp unclassified.\*  
15: sp rvirus.\*  
16: sp bacteriap.\*  
17: sp archaeap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	5184	71.6	977	6	002776 bos taurus

2	4694.5	64.8	976	4	Q86W56
3	4689.5	64.8	976	4	Q72742
4	4635.5	64.0	976	4	Q9Y4W7
5	4332.5	59.8	968	11	O88622
6	4329	59.8	961	11	Q80Y06
7	4295.5	59.3	972	11	Q9QYX2
8	4125	57.0	920	11	Q8CB72
9	1063.5	14.7	768	5	O46043
10	1061.5	14.7	723	5	Q960N8
11	585	8.1	548	10	Q9SKB3
12	533	7.4	522	10	Q8VYA1
13	481.5	6.6	781	5	Q867X0
14	475	6.6	764	5	Q19637
15	472	6.5	485	5	Q9NSL4
16	335.5	4.6	368	5	Q86G14
17	331	4.6	364	10	Q9SKB4
18	286.5	4.0	747	4	Q86V50
19	185.5	2.6	954	5	Q9VU43
20	185	2.6	508	12	Q9E234
21	185	2.6	2081	10	Q9LH98
22	181.5	2.5	1618	16	Q7UTM1
23	181	2.5	571	5	Q8MTN7
24	180	2.5	510	12	Q99GU9
25	179.5	2.5	1078	5	Q963T1
26	179	2.5	1160	11	Q9ESC8
27	177.5	2.5	1320	11	Q7TPZ2
28	175.5	2.4	1099	10	Q8W362
29	175.5	2.4	1099	10	Q7XDA1
30	175.5	2.4	1163	4	Q9UHB7
31	175.5	2.4	3942	11	O88737
32	175.5	2.4	5327	5	O76891
33	174.5	2.4	669	16	Q7UE67
34	174	2.4	3398	13	Q8QFV6
35	171.5	2.4	1444	5	Q9VTN2
36	171.5	2.4	1514	5	Q8SY55
37	171.5	2.4	1616	4	Q86PH3
38	171.5	2.4	1758	4	Q7Z6E8
39	171.5	2.4	1792	4	Q7Z6E9
40	171.5	2.4	5322	5	Q9VPL9
41	171	2.4	791	13	Q9DGL1
42	171	2.4	1183	5	Q9W3D1
43	170.5	2.4	1200	5	Q25388
44	170.5	2.4	1226	16	Q7UL12
45	170.5	2.4	1576	3	O13308

## ALIGNMENTS

RESULT 1  
002776 PRELIMINARY; PRT; 977 AA.  
ID 002776  
AC 002776;  
DT 01-JUL-1997 (TREMBlrel. 04, Created)  
DT 01-JUL-1997 (TREMBlrel. 04, Last sequence update)  
DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)  
DE Poly(ADP-Ribose) glycohydrolase.  
GN BPARG.  
OS Bos taurus (Bovine).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
OC Bovidae; Bovinae; Bos.  
OX NCBI\_taxID=9913;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=97277328; PubMed=9115250;  
RA Lin W., Ame J.C., Aboul-Ela N., Jacobson E.L., Jacobson M.K.;  
RT "Isolation and characterization of the cDNA encoding bovine poly(ADP-  
ribose) glycohydrolase."  
RL J. Biol. Chem. 272:11895-11901(1997).  
DR EMBL; U78975; AAB53370.1; -;  
DR GO; GO:0016787; F:hydrolase activity; IEA.  
DR InterPro; IPR007724; PARG.  
DR Pfam; PF05028; PARG; 1.



Db 661 AsnPheAsnArgLeuPheGluGlyArgSerArgLysProGluLysLeuLysThrLeu 680  
QY 2298 TTCTGCTACTTTAGAGAGTCACAGAGAAAACCCAGTGGTGGTGCATTCACAGA 2357  
Db 681 PheCysTyrPheArgArgValThrGluLysProThrGlyLeuValThrPheThrArg 700  
QY 2358 CAGAGTCTTGAAGATTTTCCAGAGTGGGAAGATGTCAAAATCTCCAGTCCAGTCCAT 2417  
Db 701 GlnSerLeuGluAspPheProGluTyrPgluArgCysGluLysLeuThrArgLeuHis 720  
QY 2418 GTCACTTACAGAGTACATAGAGAAACCGCCAGCGCATGTCTACAGGTGGATTTTGA 2477  
Db 721 ValThrTyrGluGlyThrIleGluGlyAsnGlyGlnGlyMetLeuGlnValAspPheAla 740  
QY 2478 AACGGTTTCCTGGAGTGGTGTAAACAGTGCAGGACTTGTGCAGAGAAATCCGCTTT 2537  
Db 741 AsnArgPheValGlyGlyGlyValThrSerAlaGlyLeuValGlnGluGluIleArgPhe 760  
QY 2538 TTAATCAACCTCAGTGTGATTGTTTTCACGGCTCTTCACTGAGGTGCTGGATCAATGAA 2597  
Db 761 LeuIleAsnProGluLeuIleValSerArgLeuPheThrGluValLeuAspHisAsnGlu 780  
QY 2598 TGTCTTATCATCAGTACTAGCAGTACAGTGAATACACAGGCTATGCCGAAACATAC 2657  
Db 781 CysLeuIleIleThrGlyThrGluGlnTyrSerGluTyrThrGlyTyrAlaGluThrTyr 800  
QY 2658 CGCTGGCCGGAGCCATGAAGACAGGAGCGAAGGACGACACTGGCAGAGCCGACGACT 2717  
Db 801 ArgTyrAlaArgSerHisGluAspArgSerGluArgAspAspTyrPgluArgThrThr 820  
QY 2718 GAGATCGTCCCATCCAGCGCCCTCCACTTCAGACGCTACCTCGACAGTGTGTGCCGAG 2777  
Db 821 GluIleValAlaIleAspAlaLeuHisPheArgArgTyrLeuAspGlnPheValProGlu 840  
QY 2778 AGATCAGACGGAGCTTAACAGGCTTACTGTGGATTCTCTGCTCGAGTTCTTCA 2837  
Db 841 LysIleArgArgGluLeuAsnLysAlaTyrCysGlyPheLeuArgProGlyValSerSer 860  
QY 2838 GAGAACTCTCTCAGTGGCTACAGAACTGGGCTGGTGGTCCCTTGGGGTGATGCT 2897  
Db 861 GluAsnLeuSerAlaValAlaThrGlyAsnTyrPgluArgGlyAlaPheGlyGlyAspAla 880  
QY 2898 AGACTAAAGCCCTTAATACAGATCTCGGACGCTGTAGTCCAGCAGACGCGGTAT 2957  
Db 881 ArgLeuLysAlaLeuIleGlnIleLeuAlaAlaValAlaGluArgAspValValTyr 900  
QY 2958 TTCACCTTTGGGACTCAGAACTGATGAGAGCATTTACAGCATGATCATTCCTCAT 3017  
Db 901 PheThrPheGlyAspSerGluLeuMetArgAspIleTyrSerMetHisThrPheLeuThr 920  
QY 3018 GAGAGAACTGACTCTTCGAGAGTATATAGCTCTGCTAGCATTTACATGAGAA 3077  
Db 921 GluArgLysLeuThrValGlyGluValTyrLysLeuLeuArgTyrTyrAsnGluGlu 940  
QY 3078 TCAGAACTGCTCCACCCCGACACAGATCAAGCTTTATCCATTCATATACATGCA 3137  
Db 941 CysArgAsnCysSerThrProGlyProAspIleLysLeuTyrProPheIleTyrHisAla 960  
QY 3138 GTTGAGTCTCTGACAGACCAACACCGCCGGGCAAGAGACGGGGCC 3188  
Db 961 ValGluSerCysThrGlnThrThrAsnGlnProGlyGlnArgThrGlyAla 977

## RESULT 2

Q86W56 PRELIMINARY; PRT; 976 AA.  
ID Q86W56;  
AC Q86W56;  
DT 01-JUN-2003 (TrEMBLrel. 24, Created)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Similar to poly (ADP-ribose) glycohydrolase.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
QX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Skin;  
RA Strausberg R.;  
RL Submitted (APR-2003) to the EMBL/GenBank/DBJ databases.  
DR EMBL: BC050560; AAH50560.1; -;  
DR GO: 0016787; F:hydrolase activity; IEA.  
DR InterPro: IPR007724; PARG.  
DR Pfam: PF05028; PARG; 1.  
KW Hydrolase.  
SQ SEQUENCE 976 AA; 111109 MW; D6646353C6D0180E CRC64;  
Alignment Scores:  
Pred. No.: 0 Length: 976  
Score: 4694.50 Matches: 883  
Percent Similarity: 93.44% Conservative: 29  
Best Local Similarity: 90.47% Mismatches: 63  
Query Match: 64.82% Indels: 1  
DB: 4 Gaps: 1

## US-09-302-812-1 (1-4070) x Q86W56 (1-976)

QY 258 ATGAGTGGCGGCGCGCTGTGAGCCCTGCACCAAGCAGCCCGCTGGAGCGCGCTGCA 317  
Db 1 MetAsnAlaGlyProGlyCysGluProCysThrLysArgProArgTyrPgluAlaAlaThr 20  
QY 318 ACTTCTCCCGCCGCGCTCCGACCGCCGAGCTCCCGCGCAGGAGGAGCGCGCTCTC 377  
Db 21 ThrSer---ProAlaAlaSerAspAlaArgSerPheProSerArgGlnArgValLeu 39  
QY 378 GATTCACAGAGCGCTCCGCTGAGTTCAGGTCCCGCGCTCCCTCTCGTCAGCTCGCGCTG 437  
Db 40 AspProLysAspAlaHisValGlnPheArgValProProSerSerProLysValPro 59  
QY 438 GGC CGCGCGGCGCAGCAGCAGCGCGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 497  
Db 60 GlyArgAlaGlyGlnHisArgGlySerAlaThrSerLeuValPheLysGlnLysThrIle 79  
QY 498 ACCAGTTGGATGGACACTAAAGGAATCAAGACAGATTTGAATCAAGAAAGTTTGCATAGTAA 557  
Db 80 ThrSerTyrMetAspThrLysGlyIleLysThrAlaGluSerGluSerLeuAspSerLys 99  
QY 558 GAAACAACAATACAGAGAGAAATCCATGATGAGTCTCTGTACAAAAGATTAATCTTTAT 617  
Db 100 GluAsnAsnAsnThrArgIleGluSerMetSerValGlnLysAspAsnPheTyr 119  
QY 618 CAACATAACATGGAATAATAGAAAATGTTTCTCAGCTAGGTTTGTATAGTCACCATGTT 677  
Db 120 GlnHisAsnValGluLysLeuGluAsnValSerGlnLeuSerLeuAspLysSerProThr 139  
QY 678 GAAAGAAGTACAGATTTTGAAGCAGCATCAGATCGCGCTATGTGTAAAGTGGCAGAAAT 737  
Db 140 GluLysSerThrGlnTyrLeuAsnGlnHisGlnThrAlaAlaMetCysLysTyrPgluAsn 159  
QY 738 GAAGGCGCACACTCAGACGCGCTTTTGAAGTGAACCTCCAGCGCTAACTCTGGTACCA 797  
Db 160 GluGlyLysHisThrGluGlnLeuLeuGluSerGluProGlnThrValThrLeuValPro 179  
QY 798 GAGCAGTTTCAGTAATGCTAATGCTCAGTCAGTCCGTCGTCGTCGTCGTCGTCGTCGTC 857  
Db 180 GluGlnPheSerAsnAlaAsnIleAspArgSerProGlnAsnAspAspHisSerAspThr 199  
QY 858 AATAGTGGAGAGTAGAGATTAATCAGCAGATTTTGTGACATGTAAAGTTCGCAATGCA 917  
Db 200 AspSerGluGluAsnArgAspAsnGlnGlnPheLeuThrThrValLysLeuAlaAsnAla 219  
QY 918 AAGCAGAGCATGAGATTCACAGCGCGCAGAGAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 977  
Db 220 LysGlnThrThrGluAspGluGlnAlaArgGluAlaLysSerHisGlnLysCysSerLys 239  
QY 978 GCTTGCCTCTCTGCAGAGCGCTGTGCGGGTGTGACAGGAGGAGGAGCAGACGCTGGTGTCC 1037

```
Db 240 SerCysAspProGlyGluAspCysAlaSerCysGlnGlnAspGluLeuAspValValPro 259
QY 1038 GAGAGCCCTGTCGACACTGGCTGAGGATGTTGGTACTGGAGTGAATAATGCCAAC 1097
Db 260 GluSerProLeuSerAspValGlySerGluAspValGlyThrGlyProLysAsnAspAsn 279
QY 1098 AGATTGAATAGACAAGAAAGTAGTCTAGGAAATCTCTCCATTTCTCCATTTGAGAAAGAAAGTGA 1157
Db 280 LysLeuThrArgGlnGlnSerCysLeuGlyAsnSerProProPheGluLysGluSerGlu 299
QY 1158 CTGAGTACCAATAGTATAGATTAATCCAAAATAGTTGTGAGATTTCAGAGCAGAT 1217
Db 300 ProGluSerProMetAspValAspAsnSerLysAsnSerCysGlnAspSerGluAlaAsp 319
QY 1218 GAAGACAGAGTCCAGGTTTTCATGAACAGGAGATAGCAGTTCCTCAACACGCAAT 1277
Db 320 GluGluThrSerProGlyPheAspGluGlnGluAspGlySerSerGlnThrAlaAsn 339
QY 1278 AAACCTTCAAGGTTCCAAACAGAGAAAGCTGACACTGAGTTGAGGAAGCGTCTCTGCT 1337
Db 340 LysProSerArgPheGlnAlaArgAspAlaAspIleGluPheArgLysArgTyrSerThr 359
QY 1338 AAGGAGGTGAGATTGCGATTACATTTCCAAATTTGAAGGAGGAGAGAGTGGAGCTGGAATG 1397
Db 360 LysGlyGlyGluValArgLeuHisPheGlnPheGluGlyGlyGluSerArgThrGlyMet 379
QY 1398 AATGATGTGAATGCCAAACGACTGGAAGTACTTCTAGCCTGAATGTAGTGCAGAAAT 1457
Db 380 AsnAspLeuAsnAlaLysLeuProGlyAsnIleSerSerLeuAsnValGluCysArgAsn 399
QY 1458 TCTAAGCAACATGGGAGAAAGGATTCTAAATTCACAGATCATTTTCATGAGAGTGCCCAA 1517
Db 400 SerLysGlnHisGlyLysLysAspSerLysIleThrAspHisPheMetArgLeuProLys 419
QY 1518 GCAGAGGCAAAAGAAAGAAACAATGTGAATGAACATCAAGAACACAGAAAGGAGATC 1577
Db 420 AlaGluAspArgLysGluGlnTrpGluThrLysHisGlnArgThrGluArgLysIle 439
QY 1578 CCTAAATACATTCACCTTCACCTTCTCCAGATAGAAATGGCTTGGAACTCCTATTGAG 1637
Db 440 ProLysTyrValProProHisLeuSerProAspLysLysTrpLeuGlyThrProLysGlu 459
QY 1638 GAGATGAGGAGAAATGCCAAGGTGTGGATCGCGTCTCCCTTGAGACCATCTGCCAAT 1697
Db 460 GluMetArgArgMetProArgCysGlyIleArgLeuProLeuLeuArgProSerAlaAsn 479
QY 1698 CACACAGTGACTATTTCGGGTAGATCTTTTGGGAATAGAGAGATTCCTAAACCTTCCCA 1757
Db 480 HisThrValThrIleArgValAspLeuLeuArgAlaGlyGluValProLysProPhePro 499
QY 1758 ACACATTTTAAAGATTCTGGCAACAACAGCATGTTAAGATGCTTGTTCAGAACAAAC 1817
Db 500 ThrHisTyrLysAspLeuTrpAspAsnLysHisValLysMetProCysSerGluGlnAsn 519
QY 1818 TTGTACCTGTGGAAGATGAGAATGGTGGAGAGCTGCAGGACGCGGTGGGAATCTATT 1877
Db 520 LeuTyrProValGluAspGluAsnGlyGluArgThrAlaGlySerArgTrpGluLeuIle 539
QY 1878 CAGACTGCACTTCTCAACAGGCTACTCGGCCCGCAGAACCTGGAAGATGCTATTCTGAAG 1937
Db 540 GlnThrAlaLeuLeuAsnLysPheThrArgProGlnAsnLeuLysAspAlaIleLeuLys 559
QY 1938 TACAATGTGGCATATTCTAAGAAATGGGACTTTACAGCTTTGATTGATTCTGGGATAAG 1997
Db 560 TyrAsnValAlaTyrSerLysLysTrpAspPheThrAlaLeuIleAspPheTrpAspLys 579
QY 1998 GTACTAGAAGAGCAGAGAGCTCAACACATTGTATCAGTCCATCTTGCCTGTATGCTGAAA 2057
Db 580 ValLeuGluGluAlaGluAlaGlnHisLeuTyrGlnSerIleLeuProAspMetValLys 599
QY 2058 ATTGCACTCTGTGCGCAAAATATTGTATCCAGCAATATACACTCTCTGGAACACAGAGATG 2117
```

---

```
600 IleAlaLeuCysLeuProAsnIleCysThrGlnProIleProLeuLeuLysGlnLysMet 619
QY 2118 AATCAATTCATCACAAATGTCTCAGAGAAACAGATTGGCAGTCTTTTATAGCTTAATGCTTTCTTC 2177
Db 620 AsnHisSerIleThrMetSerGlnGluGlnIleAlaSerLeuLeuAlaAsnAlaPhePhe 639
QY 2178 TGCAGCTTTCCACGACGCAATGCCAAGATGAATCAGAGTATTTCAGTATTATCCAGATATT 2237
Db 640 CysThrPheProArgArgAsnAlaLysMetLysSerGluTyrSerSerTyrProAspIle 659
QY 2238 AACTCAATTCGGTGTGTTTGAAGAGCTTCATCAAGGAAACACAGAGAACTTAAAAACGCTC 2297
Db 660 AsnPheAsnArgLeuPheGluGlyArgSerSerArgLysProGluLysLeuLysThrLeu 679
QY 2298 TTCTGCTACTTTAGAGAGTACAGAGAGAAACCCACTGGGTTGGTGCACATTCAACAAG 2357
Db 680 PheCysTyrPheArgArgValThrGluLysLysProThrGlyLeuValThrPheThrArg 699
QY 2358 CAGAGTCTTGAAGATTTTCCAGAGTGGGAAAGATGTGAAAACTCTCTCACTCGACTGCAT 2417
Db 700 GlnSerLeuGluAspPheProGluTrpGluArgCysGluLysProLeuThrArgLeuHis 719
QY 2418 GTCACTTACGAAGGTACCATAGAAGAAACCGCCAGGCGCATGCTACAGTGGATTTCGA 2477
Db 720 ValThrTyrGluGlyThrIleGluGluAsnGlyGlnGlyMetLeuGlnValAspPheAla 739
QY 2478 AACCGTTTCGTTGGAGGTGTGTAAACAGTGCAGGACTTGTGCAAGAGAAATCCGCTTT 2537
Db 740 AsnArgPheValGlyGlyValThrSerAlaGlyLeuValGlnGluGluIleArgPhe 759
QY 2538 TTAATCAACCTGAGTGTGTTTACGGCTCTTCACTGAGGTGCTGGATCACAATGAA 2597
Db 760 LeuLeuAsnProGluLeuIleIleSerArgLeuPheThrGluValLeuAspHisAsnGlu 779
QY 2598 TGTCTTATCATCACAGGTACTGAGCAGTACAGTGAATACACAGGCTATSCCGAAACATAC 2657
Db 780 CysLeuIleIleThrGlyThrGluGlnTyrSerGluTyrThrGlyTyrAlaGluThrTyr 799
QY 2658 CGCTGGGCCCGGAGCCATGAAGACAGAGCGAAAGGAGCAGTGCAGAGGCGGCAGCT 2717
Db 800 ArgTrpSerArgSerHisGluAspGlySerGluArgAspAspTrpGlnArgArgCysThr 819
QY 2718 GAGATCGTGGCATTCGACGCGCTCCACTTCAGAGCTACCTCGACACCACTTTGTGCGCCAG 2777
Db 820 GluIleValAlaIleAspAlaLeuHisPheArgArgTyrLeuAspGlnPheValProGlu 839
QY 2778 AAGATCAGACGGGAGCTTAACAAGCTTACTGTGGATTCTTCTGCTGGAGTTTCTTCA 2837
Db 840 LysMetArgArgGluLeuAsnLysAlaTyrCysGlyPheLeuArgProGlyValSerSer 859
QY 2838 GAGAACCTGTCTGAGTGGCTTACAGGAAACCTGGGCGCTGTGTGCTTTGGGGGTGATGCT 2897
Db 860 GluAsnLeuSerAlaValAlaThrGlyAsnTrpGlyCysGlyAlaPheGlyGlyAspAla 879
QY 2898 AGACTAAAAGCTTAATACAGATCTCTGCGAGCTGCTAGCTGAGCAGAGCTGTTTAT 2957
Db 880 ArgLeuLysAlaLeuIleGlnIleLeuAlaAlaAlaAlaGluArgAspValValTyr 899
QY 2958 TTCACTTTTGGGAGCTCAGAACTGATGAGAGACATTTACAGCATGCATATCTCTCACT 3017
Db 900 PheThrPheGlyAspSerGluLeuMetArgAspIleTyrSerMetHisIlePheLeuThr 919
QY 3018 GAGAGGAAACTGACTGTTGGAGAGATATAAGCTGCTGTCTACGATATTACATGAAGAA 3077
Db 920 GluArgLysLeuThrValGlyAspValTyrLysLeuLeuLeuArgTyrTyrAsnGluGlu 939
QY 3078 TGCAGAACTCTCCACCCCGGACACAGACATCAAGCTTTTATCCATTTCATATACCATGCA 3137
Db 940 CysArgAsnCysSerThrProGlyProAspIleLysLeuTyrProPheIleTyrHisAla 959
QY 3138 GTTGAGTCTGTGTACACAGACCAACACCGCGGACAAAGAGCGGG 3185
Db 960 ValGluSerCysAlaGluThrAlaAspHisSerGlyGlnArgThrGly 975
```



```

RESULT 3
Q7Z742 PRELIMINARY; PRT; 976 AA.
AC Q7Z742;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DE Hypothetical protein.
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Testis;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.D., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Udwin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Locuelli N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Hellon E., Kettman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzyzanski M.I., Skalska U., Smailus D.E., Scherch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Testis;
RA Strausberg R.;
RL Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC02966; AAHS2966.1; -.
KW Hypothetical protein.
SQ SEQUENCE 976 AA; 111090 MW; 47C8BE1826C4A74A CRC64;

Alignment Scores:
Pred. No.: 0 Length: 976
Score: 4689.50 Matches: 882
Percent Similarity: 93.34% Conservative: 29
Best Local Similarity: 90.37% Mismatches: 64
Query Match: 64.75% Indels: 1
DB: 4 Gaps: 1

US-09-302-812-1 (1-4070) x Q7Z742 (1-976)
QY 258 ATGAGTGGGGCCCGCGTGTGAGCCCTGCACAAAGCGACCCGCTGGAGCGCCCTGCA 317
DB 1 MetAenAlaGlyProGlyCysGluProCysThrLysArgProArgTrpGlyAlaThr 20
QY 318 ACTTCTCCGCGCGCGCTCGGACCCCGGAGCTTCCCGGCGAGCGAGCGCGCTCTC 377
DB 21 ThrSer---ProAlaAlaSerAspAlaArgSerPheProSerArgGlnArgValLeu 39
QY 378 GATTCACAGGAGCGTCCGCTGAGTTCAGGTCGCGCGCTCTCTGTCAGGCTGCGCCCTG 437
DB 40 AspProLysAspAlaHisValGlnPheArgValProProSerProAlaCysValPro 59
QY 438 GCGCCGGGGGACAGACACCGAGGCGAGCGCCACTCTCTTGTGTTTCAAACAGAGACTATA 497
DB 60 GlyArgAlaGlyGlnHisArgGlySerAlaThrSerLeuValPheLysGlnLysThrIle 79
QY 498 ACCAGTGGATGGACACTAAAGGAATCAAGACAGTGTGAATCAGAAAGTTTCAGTAGTAA 557

```

```

DB 80 ThrSerTrpMetAspThrLysGlyIleLysThrAlaGluSerGluSerLeuAspSerLys 99
QY 558 GAAACCAACATACAGAGAGAAGATCCCATGATGAGTTCGTGTAACAAAGATAACTTTTAT 617
DB 100 GluAenAsnAsnThrArgIleGluSerMetSerValGlnLysAspAsnPheTyr 119
QY 618 CAACATACATCGGAAATATGAAATCTTCTCAGCTAGGTGTTTGCATAGTCACAGTT 677
DB 120 GlnHisAsnValGluLysLeuGluAsnValSerGlnLeuSerLeuAspLysSerProThr 139
QY 678 GAAAGAGGTACACAGTATTGAGCAGCATCAGATCGCGCTATGTGAAGTGGCAGAT 737
DB 140 GluLysSerThrGlnTyrLeuAsnGlnHisGlnThrAlaAlaMetCysLysTrpGlnAsn 159
QY 738 GAAGGCCCACTCAAGACGGCTTTTGGAAAGTGAACCTCCAGCGGTAACTCTGGTACCA 797
DB 160 GluGlyLysHisThrGluGlnLeuLeuGluSerGluProGlnThrValThrLeuValPro 179
QY 798 GAGCAGTTTCAGTAAAGTAACTCAGTCAGTCGTCCCAAGAGATGATCAGATGACACA 857
DB 180 GluGlnPheSerAsnAlaAsnIleAspArgSerProGlnAsnAspAspHisSerAspThr 199
QY 858 AATAGTGGAGAGTAGAGATAATCAGCAGATTTTGGACACATGTAAAGCTTCGCAATGCA 917
DB 200 AspSerGluGluAsnArgAspAsnGlnGlnPheLeuThrThrValLysLeuAlaAsnAla 219
QY 918 AAGCAGCATGGAAGATGAACAGCGGCAGAGAGCCAGCCAGACCCAGAGTGTGGCAAG 977
DB 220 LysGlnThrThrGluAspGluAlaArgGluAlaLysSerHisGlnLysCysSerLys 239
QY 978 GCTTCCCATCTCGCAGAGACCTGTGCAGGGTGTCTCAGCAGGAGGACAGACAGCTGTGTC 1037
DB 240 SerCysAspProGlyGluAspCysAlaSerCysGlnGlnAspGluIleAspValValPro 259
QY 1038 GAGACCCCTTGTTCGACACTCGGCTCTCAGGATGTTGTGCTACTGGACTCAAAATGCCAAC 1097
DB 260 GluSerProLeuSerAspValGlySerGluAspValGlyThrGlyProLysAsnAspAsn 279
QY 1098 AGATTGAATAGACAAAGAGTAGTCTAGGAATCTCTCCATTTTGAGAGAGAAGTGA 1157
DB 280 LysLeuThrArgGlnGluSerCysLeuGlyAsnSerProProPheGluLysGluSerGlu 299
QY 1158 CTTGAGTCAACCAATGGATGATGATAATCCAAAATAGTTGTGACGATTTCAGAGCAGAT 1217
DB 300 ProGluSerProMetAspValAspAsnSerLysAsnSerCysGlnAspSerGluAlaAsp 319
QY 1218 GAAGAGCAAGTCCAGGTTTGTGATGAACAGGAAGATAGCAGTTCCTGTCACAAACAGCAAT 1277
DB 320 GluGluThrSerProGlyPheAspGluGlnGluAspGlySerSerSerGlnThrAlaAsn 339
QY 1278 AAACCTTCAAGGTTCCACCAAGAGAGCTCAGCTGAGTTGAGGAAGCGGCTCTGCT 1337
DB 340 LysProSerArgPheGlnAlaArgAspAlaAspIleGluPheArgLysArgTyrSerThr 359
QY 1338 AAGGAGGTGAGATTCGATTTCATTTTCAATTTTCAAGGAGGAGAGAGTTCGAGCTGGAGATG 1397
DB 360 LysGlyGlyGluValArgLeuHisPheGlnPheGluGlyGlyGluSerArgThrGlyMet 379
QY 1398 AATGATGTGAATGCCAACAGCAGCTGGAAGTACTTCTAGCTGAATGTAGATGAGCAAAAT 1457
DB 380 AsnAspLeuAsnAlaLysLeuProGlyAsnIleSerSerLeuAsnValGluCysArgAsn 399
QY 1458 TCTAAGCAACATGGGAGAAAGGATTCTAAATACAGATCACAGTCAATTTTCATCAGAGTCCCAA 1517
DB 400 SerLysGlnHisGlyLysLysAspSerLysIleThrAspHisPheMetArgLeuProLys 419
QY 1518 GCAGAGACAAAGAAAGAAACAAATGTGAATGAACATCAAGAAACAGAAAGCAAGATC 1577
DB 420 AlaGluAspArgArgLysGluGlnTrpGluThrLysHisGlnArgThrGluArgLysIle 439
QY 1578 CCTAAATACATTCACCTCCCTCTTCTCCAGATGAAGAAATGGCTTGGAACTCTTATTGAG 1637

```

```
Db 440 ProLysTyrValProProHisLeuSerProAspLysLysTrpLeuGlyThrProIleGlu 459
Qy 1638 GAGATGAGGAGAAATGCCAAGGTGTGGGATCCGGCTGCCTCCTTGGAGACCAATCGCAAT 1697
Db 460 GluMetArgArgMetProArgCysGlyIleArgLeuProLeuLeuLeuArgProSerAlaAsn 479
Qy 1698 CACACAGTGACTATTTCGGGTAGATCTTTTGGGAATAGAGAAAGTTCTTAAACCTTTCCCA 1757
Db 480 HisThrValThrIleArgValAspLeuLeuLeuArgAlaGlyGluValProLysProPhePro 499
Qy 1758 ACACATTTTAAAGATTTCGGGACAAACAGCATGTTAAGATGCCCTTGTTCAGAAACAAAC 1817
Db 500 ThrHisTyrLysAspLeuTyrAspAsnLysHisValLysMetProCysSerGluGlnAsn 519
Qy 1818 TTGTACCTCTGGAGATGAGAAATGGTGAGGAGCTGCAGCAGCCGGTGGGAATCATTT 1877
Db 520 LeuTyrProValGluAspGluAsnGlyGluArgThrAlaGlySerArgTrpGluLeuIle 539
Qy 1878 CAGACTGCACCTCTCAACAGAGCTCACTCGGCCCCAGAACCTGAAGGATGCTATTCTGAAG 1937
Db 540 GlnThrAlaLeuLeuAsnLysPheThrArgProGlnAsnLeuLysAspAlaIleLysLys 559
Qy 1938 TACAATGTGGCATATCTCAAGAAATGGGACTTTACAGCTTTGATGATTTCTGGGATAAG 1997
Db 560 TyrAsnValAlaTyrSerLysLysTrpAspPheThrAlaLeuIleAspPheTrpAspLys 579
Qy 1998 GTACTAGAGAGCAGAGCTCAACACTTGATCATGTCATCTTGCCTGATATGTTGAAA 2057
Db 580 ValLeuGluGluAlaGluAlaGlnHisLeuTyrGlnSerIleLeuProAspMetValLys 599
Qy 2058 ATTGCACCTCTCTGCCAAATATTGTACCCAGCAATATACACTCTCTGAAACAGAGATG 2117
Db 600 IleAlaLeuCysLeuProAsnIleCysThrGlnProIleProLeuLeuLysGlnLysMet 619
Qy 2118 AATCATTTCCATCAATGTCCAGAACAGATTCAGCTCTTTTGTAGCTAATGCTTTCTTC 2177
Db 620 AsnHisSerIleThrMetSerGlnGluGlnIleAlaSerLeuLeuAlaAsnAlaPhePhe 639
Qy 2178 TGCAGCTTCCACAGCAATGCCAAGATGAAATCAGAGTATTCAGTATTCAGATATTT 2237
Db 640 CysThrPheProArgAsnAlaLysMetLysSerGluTyrSerSerTyrProAspIle 659
Qy 2238 AACTTCAATCGGTTGTTGAAGGAGCTTCATCAAGAAACAGAGAGCTTAAACGCTC 2297
Db 660 AsnPheAsnArgLeuPheGluGlyArgSerSerArgLysProGluLysLeuLysThrLeu 679
Qy 2298 TTCTGCTACTTTAGAGGTACAGAGAACAAACCCAGCTGGTGGTGACATTCACAGA 2357
Db 680 PheCysTyrPheArgValThrGluLysLysProThrGlyLeuValThrPheThrArg 699
Qy 2358 CAGAGCTTGAAGATTTTCCAGAGCTGGGAAAGATGTGAAGAACTCCTGACTCGACTGCAT 2417
Db 700 GlnSerLeuGluAspPheProGluTrpGluArgCysGluLysProLeuThrArgLeuHis 719
Qy 2418 GTCACTTACGAAGTACCATAGAAGGAAACGCCAGGCGCTGTACAGGTGATTTTGCA 2477
Db 720 ValThrTyrGluGlyThrIleGluGluAsnGlyGlnGlyMetLeuGlnValAspPheAla 739
Qy 2478 AACCGTTTCGTGGAGTGTGTAAACAGTCCAGACTTGTGCAAGAGAAATCCGCTTT 2537
Db 740 AsnArgPheValGlyGlyValThrSerAlaGlyLeuValGlnGluIleArgPhe 759
Qy 2538 TTAATCAACCTGAGTTGATTTTCAGGCTCTTCACTGAGGTGCTGGATCAACATCAA 2597
Db 760 LeuIleAsnProGluLeuIleIleSerArgLeuPheThrGluValLeuAspHisAsnGlu 779
Qy 2598 TGTCTTATCATCAGAGTACTGAGCAGTACAGTGAATACACAGGCTATCGCGAAACATAC 2657
Db 780 CysLeuIleIleThrGlyThrGluGlnTyrSerGluTyrThrGlyTyrAlaGluThrTyr 799
Qy 2658 CGCTGGGCCCGGAGCATGAGAGCAGAGCGGAAAGGAGCGACTCGCAGAGCGCGCACT 2717
Db 800 ArgTrpSerArgSerHisGluAspGlySerGluArgAspTrpGlnArgHisCysThr 819
```

```
Qy 2718 GAGATCTCTGCCATCGACGCCCTCCACTTCAGAGCTACCTCGACAGTTTGTGCCGAG 2777
Db 820 GluIleValAlaIleAspAlaLeuHisPheArgArgTyrLeuAspGlnPheValProGlu 839
Qy 2778 AAGATCAGACGGGAGCTTAACAAGGCTTACTGTGGATTCTTCGTCTCTCGAGTTTCTTCA 2837
Db 840 LysMetArgArgGluLeuAsnLysAlaTyrCysGlyPheLeuArgProGlyValSerSer 859
Qy 2838 GAGAACCTGTCTGAGTGGCTACAGGAAACTGGGGCTGTGGTCTTTGGGGGTGATGCT 2897
Db 860 GluAsnLeuSerAlaValAlaThrGlyAsnTrpGlyCysGlyAlaPheGlyGlyAspAla 879
Qy 2898 AGACTAAAGCCCTAATACAGATCCTCGCAGCTCTGTAGCTGACGAGCGTGTGTTAT 2957
Db 880 ArgLeuLysAlaLeuIleGlnIleLeuAlaAlaAlaGluArgAspValValTyr 899
Qy 2958 TTCACCTTTGGGACTCAGAACTGATGAGAGACATTTACAGCATGCATATCTCTCACT 3017
Db 900 PheThrPheGlyAspSerGluLeuMetArgAspIleTyrSerMetHisIlePheLeuThr 919
Qy 3018 GAGAGAACTGACTGTGTGGAGAAGTATTAAGCTGCTGTACGATATTACATTAAGAA 3077
Db 920 GluArgLysLeuThrValGlyAspValTyrLysLeuLeuLeuArgTyrTyrAsnGluGlu 939
Qy 3078 TGCAGAACTGCTCCACCCCGGACCAAGACATCAAGCTTTATCCATTCATATACCATGCA 3137
Db 940 CysArgAsnCysSerThrProGlyProAspLysLeuLysLeuTyrProPheIleTyrHisAla 959
Qy 3138 GTTGAGTCTGTACACAGACCAACACGACCGGGACAAAGAGCGGG 3185
Db 960 ValGluSerCysAlaGluThrAlaAspHisSerGlyGlnArgThrGly 975
```

## RESULT 4

```
Q9Y4W7 PRELIMINARY; PRT; 976 AA.
ID Q9Y4W7
AC Q9Y4W7
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Poly(ADP-ribose) glycohydrolase.
GN HPARG.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OC NCBI_TaxID=9606;
RX [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93380098; PubMed=10449915;
RA Ame J.C., Apiau F., Jacobson E.L., Jacobson M.K.;
RT "Assignment of the poly(ADP-ribose) glycohydrolase gene (PARG) to
RT human chromosome 10q11.23 and mouse chromosome 14B by in situ
RT hybridization."
RL Cytogenet. Cell Genet. 85:269-270 (1999).
DR EMBL; AF005043; AAB61614.1; -
DR Genbank; HGNC:8605; PARG.
DR GO; GO:0005737; C:cvtolasm; TAS.
DR GO; GO:0004649; F:poly(ADP-ribose) glycohydrolase activity; TAS.
DR InterPro; IPR007724; PARG.
DR Pfam; PF05028; PARG; 1.
KW Hydrolase.
SQ SEQUENCE 976 AA; 110903 MW; 42A0F51466FCE7DC CRC64;
```

## Alignment Scores:

Pred. No.:	0	Length:	976
Score:	4635.50	Matches:	874
Percent Similarity:	92.83%	Conservative:	32
Best Local Similarity:	89.55%	Mismatches:	69
Query Match:	64.01%	Indels:	1
DB:	4	Gaps:	1

US-09-302-812-1 (1-4070) x Q9Y4W7 (1-976)

258 ATGAGTCCGGCCCGCTGAGCCTGACCAAGCCGCTGCA 317  
1 MetAsnAlaGlyProGlyCysGluProCysThrLysAlaThrArgTrpGlyAlaAlaThr 20  
318 ACTTCTCCGGCCCGCTCGGAGCCGCGGAGCTTCCCGGCGCAGGAGCGCGTCTCTC 377  
21 ThrSer---ProAlaAlaSerAspAlaArgSerPheProSerArgGlnArgValLeu 39  
378 GATTCCAGACGCTCCGGTGCAGTTCAGGGTCCCGCTGCTCGTCCAGGCTCGCCCTG 437  
40 AspProLysAspAlaHisValGlnPheArgValProProSerSerProAlaCysValPro 59  
438 GCGCGGCGGCGCAGACCCAGCCAGCCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 497  
60 GlyGlnAlaGlyGlnHisArgGlySerAlaThrSerLeuValPheLysGlnLysThrIle 79  
498 ACCAGTTGGATGACACTAAAGGAATCAACACAGTGAATCAGAAAGTTTGCATAGTAA 557  
80 ThrSerTrpMetAspThrLysGlyIleLysThrAlaGluSerGluSerLeuAspSerLys 99  
558 GAAACACACATACAGAGAGAAATCCATGATGAGTTCGTATCAAAAGATATCTTTAT 617  
100 GluAsnAsnThrArgIleGluSerMetMetSerSerValGlnLysAspAsnPheTyr 119  
618 CAACATAACATGAAATAGAAATGTTCTCAGCTAGGTTTTCATAGTCCACAGTT 677  
120 GlnHisAsnValGlnLysLeuValAsnValSerGlnLeuSerLeuAspLysSerLeuThr 139  
678 GAAAGAGGTACACAGTATTTGAAGCAGCATCAGACTCGCGCTATGTGTAGTGGCAGAAT 737  
140 GluLysSerThrGlnLysLeuAsnGlnHisGlnThrAlaAlaMetCysLysTrpGlnAsn 159  
738 GAAGGCCACACTCAGACCGCTTTGGAAAGTGAACCTCCAGCGGTAACTCTGGTACCA 797  
160 GluGlyLysHisThrGluGlnLeuLeuGluSerGluProGlnThrValThrLeuValPro 179  
798 GACAGTTGAGTAAATGCTAATGTCAGTACAGTCTCCCAAGAGTATCAGCTGACACA 857  
180 GluGlnPheSerAsnAlaAsnIleAspArgSerProGlnAsnAspAspHisSerAspThr 199  
858 AATAGTGGAGGATAGATATATCAGCAGTATTTTGACACATATTAAGCTTGGCAATGCA 917  
200 AspSerGluGluAsnArgAspAsnGlnPheLeuThrThrValLysLeuAlaAsnAla 219  
918 AGCAGACCATGGAACATGACAGGGCAGAGAGCCAGAGCCACAGAGTGTGGCAAG 977  
220 LysGlnThrThrGluAspGluHisAlaArgGluAlaLysSerHisGlnLysCysSerLys 239  
978 GCTTGCCATCTGCAGAGCCTGTCAGGGTGTGACGAGGAGGAGACAGCTGGTGTCC 1037  
240 SerCysHisProGlyGluAspCysAlaSerCysGlnGlnAspGluLeuAspValPro 259  
1038 GAGAGCCCTTGTGACACTGCTCTGAGGATGTTGTAAGTCTGCACTGAAATGCAAC 1097  
260 LysSerProLeuSerAspValGlySerGluAspValGlyThrGlySerLysAsnAspAsn 279  
1098 AGATTGAATAGACAGAAAGTAGTCTAGGAATCTCTCCATTTGAGAAAGAAAGTGA 1157  
280 LysLeuIleArgGlnGluSerCysLeuGlyAsnSerProPheGluLysGluSerGlu 299  
1158 CTGAGTCCACCATGATGATAGTATATCCAAAATAGTTGTGAGGATTCAGAGCAGAT 1217  
300 ProGluSerProMetAspValAspAsnSerLysAsnSerCysGlnAspSerGluAlaAsp 319  
1218 GAAGAGACAAGTCCAGGTTTGTATGAACAGAGAGATAGCAGTTCCTCTCAACAGCAAT 1277  
320 GluGluThrSerProGlyPheAspGluGlnGluAspGlySerSerSerGlnThrAlaAsn 339  
1278 AAACCTTCAAGGTTCCACAGAGAGAGCTGACATGAGTGTAGGAGCGGCTCTGCT 1337  
340 LysProSerArgPheGlnAlaArgAspAlaAspIleGluPheArgLysArgTyrSerThr 359  
1338 AAGGGAGGTGAGATTCGATTACATTTCCAAATTTGAAGGAGGAGAGTCCGAGTGAATG 1397

360 LysGlyGlyGluValArgLeuLeuHisPheGlnPheGluGlyGluSerArgThrGlyMet 379  
1398 AATGATGTGAATCCCAACACGACCTGGAAGTACTTCTAGCCTGAATGTAGAGTGCAGAAAT 1457  
380 AsnAspLeuAsnAlaLysLeuProGlyAsnIleSerSerLeuAsnValGluCysArgAsn 399  
1458 TCTAAGCAACATGGGAGAAAGGATTTCTAAATACAGATCATTTCTAGAGTGCACAAA 1517  
400 SerLysGlnHisGlyLysLysAspSerLysIleThrAspHisLeuMetArgLeuProLys 419  
1518 GCAGAGCAGCAAAAGAAAGCAATCTGAAATCAAAATCAAAATCAAAAGCAAGCAATC 1577  
420 AlaGluAspArgLysGluGlnTrpGluThrLysHisGlnArgThrGluArgLysIle 439  
1578 CTTAAATACATTTCCACTTCTCCAGATTAAGAAATGGCTTGGAACTCTATTAG 1637  
440 ProLysTyrValProProHisLeuSerProAspLysLysTrpLeuGlyThrProLysGlu 459  
1638 GAGATCAGAGAAATCCAAAGTGTGGATCCGCTCGCTCCCTTGAGACCATCTGCAAT 1697  
460 GluMetArgArgMetProArgCysGlyIleArgLeuProLeuLeuArgProSerAlaAsn 479  
1698 CACACAGTCACTATTCGGCTAGATCTTTTCCGAAATAGGAGAGTTCCTAAACCTTTCCCA 1757  
480 HisThrValThrIleArgValAspLeuLeuArgAlaGlyGluValProLysProPhePro 499  
1758 ACACATTTAAAGATTTGTGGACCAACAGCATGTTAAAGTGCCTTGTTCAGAACCAAAAC 1817  
500 ThrHisTyrLysAspLeuTrpAspAsnLysHisValLysMetProCysSerGluGlnAsn 519  
1818 TTGTACCTCTGCGAAGATGAGATGGTGTGAGCGAGTGTGAGCGAGCGCGTGGGAACATCATT 1877  
520 LeuTyrProValGluAspGluAsnGlyGluArgThrAlaGlySerArgTrpGluLeuIle 539  
1878 CAGACTGCACTTCTCAACAGGCTCACTCGGCCCCAGAACCTGAGAGTCTTCTGAGAG 1937  
540 GlnThrAlaLeuLeuAsnLysPheThrArgProGlnAsnLeuLysAspAlaIleLys 559  
1938 TACAATGTGGCATTTCTAAGAAATGGGACTTTTACAGCTTTTGTGATTCTGGGATAAG 1997  
560 TyrAsnValAlaTyrSerLysLysTrpAspPheThrAlaLeuIleAspPheTrpAspLys 579  
1998 GTACTAGAAAGACAGAGACTCAACACTTGTATCAGTCCATCTTGCTGATATGTTGAA 2057  
580 ValLeuGluGluAlaGluAlaGlnHisLeuTyrGlnSerIleLeuProAspMetValLys 599  
2058 ATTGCACCTCTGCTGCCAATATTTGTACCCAGCCAATACCACTCTCTGAAACAGAGATG 2117  
600 IleAlaLeuCysLeuProAsnIleCysThrGlnProIleProLeuLeuLysGlnLysMet 619  
2118 AATCATTTCCATCACAATGTCAAGAAACAGATTTGCCAGTCTTTTAGCTAATGTTCTTTC 2177  
620 AsnHisSerIleThrMetSerGlnGlnIleAlaSerLeuLeuAlaAsnAlaPhePhe 639  
2178 TCCAGTTTCCAGCGCAATCCCAAGATGAATCAGAGTATTCAGTATTCAGATAT 2237  
640 CysThrPheProArgAsnAlaLysMetLysSerGluTyrSerSerTyrProAspIle 659  
2238 AACTTCAATCGCTGTTTGAAGGAGCTTCATCAAGAAACAGAGAGCTTAAACGCTC 2297  
660 AsnPheAsnArgLeuPheGluGlyArgSerSerArgLysProGluLysLeuLysThrLeu 679  
2298 TTCTGCTACTTTAGAGAGTCAAGAGAAACCAACCTGGGTTGGTGCATTCACAGA 2357  
680 PheCysTyrPheArgArgValThrGluLysLysProThrGlyLeuValThrPheThrArg 699  
2358 CAGAGTCTTGAAGATTTCCAGAGTGGGAAAGATGTGAAAACCTCTGACTGCTGAT 2417  
700 GlnSerLeuGluAspPheProGluTrpGluArgCysGluLysProLeuThrArgLeuHis 719  
2418 GTCACTTACAGAGGTACCATGAAGAAACGCGCAGGCGATCTACAGTGGATTTCGA 2477

Db 720 ValThrTyrGluGlyThrIleGluGluAenGlyGlnGlyMetLeuGlnValAspPheAla 739  
QY 2478 AACCGTTTCGTGGAGTGGTGTAAACAGTGCAGAGACTTGTGCAAGAAATACCGCTTT 2537  
Db 740 AsnArgPheValGlyGlyValThrSerAlaGlyLeuValGlnGluIleArgPhe 759  
QY 2538 TTAATCAACCCCTGAGTGTATGTTTTCACGGCTCTTCACAGTGTGATGATCAATGAA 2597  
Db 760 LeuIleAsnProGluLeuIleIleSerArgLeuPheThrGluValLeuAspHisAsnGlu 779  
QY 2598 TGTCTTATCATCACAGTACTGAGAGTACATGATACACAGGCTATCCCAACATAC 2657  
Db 780 CysLeuIleIleThrGlyThrGluGlnTyrSerGluTyrThrGlyTyrAlaGluThrTyr 799  
QY 2658 CGTGGCCCGGAGCCATGAGACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2717  
Db 800 ArgTyrSerArgSerHisGluAspGlySerGluArgAspAspCysGluArgArgCysThr 819  
QY 2718 GAGATCGTCCCATGACGCCCTCCATCTTCAGACGCTACCTCGACAGTTTGTGCCGAG 2777  
Db 820 GluIleValAlaIleAspAlaLeuHisPheArgArgTyrLeuAspGlnPheValProGlu 839  
QY 2778 AAGATCAGACGGAGCTTAACAAGGCTTACTGTGATTTCTTCGCTGGAGTTCTTCA 2837  
Db 840 LysMetArgArgGluLeuAsnLysAlaTyrCysGlyPheLeuArgProGlyValSerSer 859  
QY 2838 GAGAACCTGTCTGAGTGGCTACAGGAACTGGGCTGTGGCTTTGGGCTGATGCT 2897  
Db 860 GluAsnLeuSerAlaValAlaThrGlyAsnTyrGlyCysGlyAlaPheGlyAspAla 879  
QY 2898 AGACTAAAGCTTAATACAGTCTCGAGTCTGTGAGTCTGTGAGTCTGTGAGTCTGTGAG 2957  
Db 880 ArgLeuLysAlaLeuIleGlnIleLeuAlaAlaAlaAlaGluArgAspValValTyr 899  
QY 2958 TTCACCTTTGGGAGCTCAGAACTGATGAGACATTTACAGCATTCATACATTCCTCACT 3017  
Db 900 PheThrPheGlyAspSerGluLeuMetArgAspIleTyrSerMetHisIlePheLeuThr 919  
QY 3018 GAGAGGAACTGACTGTGGAGAAATATAGTCTGTGCTGCTACGATATTAATCAATGAAGA 3077  
Db 920 GluArgLysLeuThrValGlyAspValTyrLysLeuLeuLeuArgTyrTyrAsnGluGlu 939  
QY 3078 TGCAGAACTGCTCACCCCGGACCAACATCAGCTTTCATTCATTCATATACATGCA 3137  
Db 940 CysArgAsnCysSerThrProGlyProAspIleLysLeuTyrProPheIleTyrHisAla 959  
QY 3138 GTTGAAGTCTGTACACAGACCAACCAACGCGGAGCAAGGAGGAGGAGGAGGAGGAGGAG 3185  
Db 960 ValGluSerCysAlaGluThrAlaAspHisSerGlyGlnArgThrGly 975

## RESULT 5

O88622 PRELIMINARY; PRT; 968 AA.  
AC O88622;  
DT 01-NOV-1998 (TrEMBLrel. 08, Created)  
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
DE Poly(ADP-ribose) glycohydrolase.  
GN PARG.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sclerognathi; Muridae; Murinae; Mus.  
CX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=99380096; PubMed=10449915;  
RA Ame J.C., Apicou F., Jacobsen E.L., Jacobson M.K.;  
RT "Assignment of the poly(ADP-ribose) glycohydrolase gene (PARG) to  
RT human chromosome 10q11.23 and mouse chromosome 14B by in situ  
RT hybridization."  
RL Cytogenet. Cell Genet. 85:269-270(1999).  
DR EMBL; AF079557; AAC28735.1; --  
DR MGD; MGI:1347094; Parg.

DR GO:0016787; F:hydrolase activity; IEA.  
DR InterPro: IPR007724; PARG.  
DR Pfam: PF05028; PARG; 1.  
KW Hydrolase. 968 AA; 109169 MW; 3981DA4809E1E007 CRC64;  
SQ SEQUENCE

## Alignment Scores:

Pred. No.: 0 Length: 968  
Score: 4332.50 Matches: 820  
Percent Similarity: 89.86% Conservative: 57  
Best Local Similarity: 84.02% Mismatches: 90  
Query Match: 59.82% Indels: 9  
DB: 11 Gaps: 5

US-09-302-812-1 (1-4070) x 088622 (1-968)

QY 258 ATGAGTGGGGCCCGCTGTGAGCCCTGCACCAAGCGACCCCGTGGAGACCCCGCTGCA 317  
Db 1 MetSerAlaGlyProGlyTyrGluProCysThrLys---AlaArgTyrGlyAlaAlaGly 19  
QY 318 ACTTCTCCCGCCGCTCGAGCGCCGAGCTTCCCGGCGAGCGAGCGAGCGAGCGAGCGAG 377  
Db 20 ThrSerAlaProThrAlaSerAspSerArgSerPheProGlyArgGlnArgValLeu 39  
QY 378 GATTCCAGGACGCTCCGCTGAGTTCAGGGTCCCGCGCTCTCTGCTCAGGGTCCGCCCTG 437  
Db 40 AspProLysAspAlaProValGlnPheArgValProProSerSerProAlaCysValSer 59  
QY 438 GCGCGGGGCGGACAGCAGCCGAGGAGCGCCACCTCTCTGTTTTCACAAAGAGACTATA 497  
Db 60 GlyArgAlaGlyProHisArgGlyAsnAlaThrSerPheValPheLysGlnLysThrIle 79  
QY 498 ACCAGTTGGATGGACACTAAAGGAATCAAGACAGTTGAATCAGAAAGTTTCATAGTAAA 557  
Db 80 ThrThrTyrMetAspThrLysGlyProLysThrAlaGluSerGlu-----SerLys 96  
QY 558 GAAACCAACATACAGAGAGAAATCCATGATGAGTTCTGTACAAAAGATAACTTTTAT 617  
Db 97 GluAsnAsnAsnThrArgIleAspSerMetSerValGlnLysAspAsnPheTyr 116  
QY 618 CAACATACATGGAAAATTAGAAAATGTTCTCAGCTAGGTTTGTGATGATCAGCAGTT 677  
Db 117 ProHisLysValGluLysLeuGluAsnValProGlnLeuAsnLeuAspLysSerProThr 136  
QY 678 GAAAGAGTACACAGTATTTGAAGCAGCATCAGACTCGGCTAGTGTAGTGTAGTGTAGT 737  
Db 137 GluLysSerSerGlnTyrLeuAsnGlnGlnThrAlaSerValCysLysTyrGlnAsn 156  
QY 738 GAAGGGCCACATCAGAACCGCTTTTGGAAAGTGAACCTCCAGCGGTAACTCTGGTACCA 797  
Db 157 GluGlyLysHisAlaGluGlnLeuLeuAlaSerGluProProAlaGlyThrProLeuPro 176  
QY 798 GAGCAGTTCAAGTAAATGCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 857  
Db 177 LysGlnLeuSerAsnAlaAsnIleGlyGlnSerProHisThrAspAspHisSerAspThr 196  
QY 858 AATAGTCAGGAGTAGAGATAATCAGCAGCTTTTGTACACATGTAAAGCTTGGAAATGCA 917  
Db 197 AspHisGluGluAspArgAspAsnGlnGlnPheLeuThrProIleLysLeuAlaAsnThr 216  
QY 918 AAGCAGACGATGAAGATGAACAGGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 977  
Db 217 LysProThrValGlyAspGlyGln-----AlaArgSerAsnCysLysCysSerGly 233  
QY 978 GCTTGGCATCTGCAGAGGCTGTGAGGGTGTGAGGGAGAGAGAGAGAGAGAGAGAGAGAG 1037  
Db 234 SerArgGlnSerValLysAspCysThrGlyCysGlnGlnGluGluValAspValLeuPro 253  
QY 1038 GAGAGCCCTTGTGCGACACTGGCTCTGAGGATGTTGGTACTGAGCTGAGAAATGCCAAC 1097  
Db 254 GluSerProLeuSerAspValGlyAlaGluAspIleGlyThrGlyProLysAsnAspAsn 273  
QY 1098 AGATTGAATAGACAGAAAGTAGTCTAGGAAATTCCTCCATTTCAGAAAGAAAGTGA 1157



Thu May 27 09:55:53 2004

DE Poly (ADP-ribose) glycohydrolase.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6; TISSUE=Brain;  
 RA Strausberg R.;  
 EL Submitted (APR-2003) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; BC050892; AAH50892.1; -  
 DR GO; GO:0016787; F:hydrolase activity; IEA.  
 DR InterPro; IPR007724; PARG.  
 DR Pfam; PF05028; PARG; 1.  
 KW Hydrolase.  
 SQ SEQUENCE 961 AA; 108580 MW; 1558F120C367CF63 CRC64;

Alignment Scores:  
 Pred. No.: 0 Length: 961  
 Score: 4329.00 Matches: 918  
 Percent Similarity: 90.09% Conservative: 55  
 Best Local Similarity: 84.42% Mismatches: 88  
 Query Match: 59.78% Indels: 8  
 DB: 11 Gaps: 4

US-09-302-812-1 (1-4070) x Q80Y06 (1-961)

QY 258 ATGAGTGGCGGCGCGGTGTGAGCCCTGCACCAAGCGACCCCGTGGAGCCCGCTGCA 317  
 Db 1 MetSerAlaGlyProGlyTrpGluProCysThrLysArgProArgTrpGlyAlaAlaGly 20  
 QY 318 ACTTCTCCGCGCGCGCTCGAGCGCCCGGAGCTTCCCGCGGAGCGAGCGCGCTCTC 377  
 Db 21 ThrSerAlaProThrAlaSerAspSerArgSerPheProGlyArgGlnArgValLeu 40  
 QY 378 GATTCCAAGCAGCTCCGCTGAGTTCAGGTCCTCCGCTCTCTGTCGAGCTGCGCCCTG 437  
 Db 41 AspProLysAlaProValGlnPheArgValProProSerSerProAlaCysValSer 60  
 QY 438 GCGCGGCGGCGAGCAGCAGCGCGCGCCGCTCTCTCTTTTCAAAAGAGACTATA 497  
 Db 61 GlyArgAlaGlyProHisArgGlyAsnAlaThrSerPheValPheLysGlnLysThrIle 80  
 QY 498 ACCGTTGGATGACACTAAGAAATCAAGAGTTCAGTTCAGTTCAGTTCAGTTCAGT 557  
 Db 81 ThrThrTrpMetAspThrLysGlyProLysThrAlaGluSerGlu-----SerLys 97  
 QY 558 GAAACAAACATACAGAGAGAGATCCATGATGATGATGATGATGATGATGATGATGAT 617  
 Db 98 GluAsnAsnAsnThrArgLysAspSerMetSerSerValGlnLysAspAsnPheTyr 117  
 QY 618 CAACATAACATGAAATAATAGAAATTTCTCAGCTAGTGTTCAGCTAGTGTTCAGCT 677  
 Db 118 ProHisLysValGluLysLeuGluAsnValProGlnLeuAsnLysAspLysSerProThr 137  
 QY 678 GAAAAAGGTACAGATATTGAGCAGCATCAGCTGCGCTATGTTAGTGCAGAGAT 737  
 Db 138 GluLysSerSerGlnLysLeuAsnGlnGlnThrAlaSerValCysLysTrpGlnAsn 157  
 QY 738 GAAGGCCACACACAGACCGCTTTTGAAGAGTGAACCTCCAGCGGTAACTCTGGTACCA 797  
 Db 158 GluGlyLysHisAlaGluGlnLeuLeuAlaSerGluProAlaGlyThrProLeuPro 177  
 QY 798 GAGCAGTTCAGTAATGTAATGTAATGTAATGTAATGTAATGTAATGTAATGTAATGTA 857  
 Db 178 LysGlnLeuSerAsnAlaAsnLysGlnSerProHisThrAspAspHisSerAspThr 197  
 QY 858 AATAGTGAGAGAGTACAGATATATCAGCATTTTTCAGCATTTTTCAGCATTTTTCAG 917  
 Db 198 AspHisGluGluAspArgAspAsnGlnPheLeuThrProLysLeuAlaAsnThr 217  
 QY 918 AAGCAGACGATGGAAGATGAACAGGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 977

218 LysProThrValGlyAspGlyGln-----AlaArgSerAsnCysLysCysSerGly 234  
 QY 978 GCTTGCCATCTCGCAGAGAGCTGTGAGGCTGTCAGCAGAGAGAGAGAGAGAGAGAGAG 1037  
 Db 235 SerArgGlnSerValLysAspCysThrGlyCysGlnGlnGlnGlnGlnGlnGlnGlnGln 254  
 QY 1038 GAGAGCCCTTGTGGGACACTGGCTCTGAGGATGTTGGTACTGAGCTGAAAAATGCCAAC 1097  
 Db 255 GluSerProLeuSerAspValGlyAlaGluAspIleGlyThrGlyProLysAsnAspAsn 274  
 QY 1098 AGATTGAATAGACAAGAAAGTAGCTAGGAAATCTCTCTCATTTGAGAAAGAAAGTCAA 1157  
 Db 275 LysLeuThrGlyGlnGluSerSerLeuGlyAspSerProProPheGluLysGluSerGlu 294  
 QY 1158 CCTCAGTCACCAATGATGTAGATAATTCAAAAATAGTTCTCAGGATTCAGAGACAGAT 1217  
 Db 295 ProGluSerProMetAspValAspAsnSerLysAsnSerCysGlnAspSerGluAlaAsp 314  
 QY 1218 GAAGAGACAAGTCCAGGTTTGTGATGAACAGAGAAATAGCAGTTCCTCAACAGCAAT 1277  
 Db 315 GluGluThrSerProValPheAspGluGlnAspAspArgSerSer--GlnThrAlaAsn 333  
 QY 1278 AAACCTTCAAGTTCACCAACAGAGAGAGCTGACACTGAGTTGAGGAGCGCTCTCTGCT 1337  
 Db 334 LysLeuSerSerCysGlnAlaArgGluAlaAspGlyAspLeuArgLysArgTyrLeuThr 353  
 QY 1338 AAGGAGGTGAGATTTCGATTACATTTCCAATTTGAAGGAGGAGAGAGTTCGAGCTGGAATG 1397  
 Db 354 LysGlySerGluValArgLeuHisPheGlnPheGlu---GlyGluAsnAsnValGlyThr 372  
 QY 1398 AATGATGTAATGCAACAGACCTGGAAAGTACTTACCTCGAATGTAGAGTGCAGAAAT 1457  
 Db 373 SerAspLeuAsnAlaLysProSerGlyAsnSerSerSerLeuAsnValGluCysArgSer 392  
 QY 1458 TCTAAGCAACATGCGGAGAAAGATTCTAAATTCACAGATCATTTTCATGAGTGCACAAA 1517  
 Db 393 SerLysGlnHisGlyLysArgAspSerLysIleThrAspHisPheMetArgLysLys 412  
 QY 1518 GCAGAGGACAAAGAAAAAGAACAAATGTGAAATGAACATCAAGAACAGAAAGAGATC 1577  
 Db 413 SerGluAspArgArgLysGluGlnCysGluValArgHisGlnArgThrGluArgLysIle 432  
 QY 1578 CCTAATAATCATTCACCTTCCTCCAGATAAGAAATGCGCTTGGAACTCCTATTGAG 1637  
 Db 433 ProLysTyrIleProProAsnLeuProGluLysLysLysLysLysLysLysLysLys 452  
 QY 1638 GAGATGAGAGATGCCAGGTGCGGATCCGCTGCTCCCTTCAGACCATCTGCAAT 1697  
 Db 453 GluMetArgLysMetProArgCysGlyIleHisLeuProSerLeuArgProSerAlaSer 472  
 QY 1698 CACACAGTCACTATTTCGGGTAGATCTTTTTCGAATAGGAGAGAGTTCCTAAACCTTTCCCA 1757  
 Db 473 HisThrValThrValArgValAspLeuLeuArgAlaGlyGluValProLysProPhePro 492  
 QY 1758 ACACATTTTAAAGATTGTTGGGACAAACAGCATGTTAAGATGCTTGTTCAGAACAAAC 1817  
 Db 493 ThrHisTyrLysAspLeuTrpAspAsnLysHisValLysMetProCysSerGluGlnAsn 512  
 QY 1818 TTGTACCTCTGGAAGATGAGATGTCGAGCAGCTGCGAGCGCGCGGTGGGAACCTCAT 1877  
 Db 513 LeuTyrProValGluAspGluAsnGlyGluArgThrAlaGlySerArgTrpGluLeuIle 532  
 QY 1878 CAGACTGCACTTCTCAACAGGCTCACTCGGCCCCAGAACCTGAGAGATGCTATTCTGAAG 1937  
 Db 533 GlnThrAlaLeuLeuAsnLysPheThrArgProGlnAsnLeuLysAspAlaIleLeuLys 552  
 QY 1938 TACAATGTGGCATATTCTAAGAAATGGGACTTTACAGCTTTCATTCATTCTTCGGATAAG 1997  
 Db 553 TyrAsnValAlaTyrSerLysLysTrpAspPheThrAlaLeuValAspPheTrpAspLys 572  
 QY 1998 GTACTAGAGAGACAGAGCTCAACACTTGTATCAGTCCATCTTGCCTGATATGATGGTGA 2057  
 Db 573 ValLeuGluGluAlaGluAlaGlnHisLeuTyrGlnSerIleLeuProAspMetValLys 592

QY 2058 ATTGACCTCTGCTGCAATATTTGTATCCAGCAATATACCACTCTCTGAAACAGAGATG 2117  
 Db 593 TleAlaLeuCysLeuProAsnileCysThrGlnProIleProLeuLeuLysGlnLysMet 612  
 QY 2118 ATATCATTCATCAGATGTCAGAGAACAGATGTCAGTCTTTTGTAGCTAATCTTCTTC 2177  
 Db 613 AsnHisSerValThrMetSerGlnGlnIleAlaSerLeuLeuAlaHisPhePhe 632  
 QY 2178 TGCACCTTTCCAGCAGCAATGCAAGATGAAATCAGAGATATTCAGATATTCAGATAT 2237  
 Db 633 CysThrPheProArgAsnAlaLysMetLysSerGluTyrSerSerTyrProAspIle 652  
 QY 2238 AACTCAATCGGTGTTTGAAGACGCTTCATCAGAAACAGAGAGCTTAAACGCTC 2297  
 Db 653 AsnPheAsnArgLeuPheGluGlyArgSerArgLysProGluLysLeuLysThrLeu 672  
 QY 2298 TTCTGCTACTTTAGAGAGCTCAGAGAAACACCACTGGTGTGTGACATTCACAGA 2357  
 Db 673 PheCysTyrPheArgArgValThrGluLysLysProThrGlyLeuValThrPheThrArg 692  
 QY 2358 CAGAGCTTTGAAGATTTCCAGAGTCGGAAGATGTGAAATCTCCTGACTCGACTGCAT 2417  
 Db 693 GlnSerLeuGluAspPheProGluTyrGluArgCysGluLysProLeuThrArgLeuHis 712  
 QY 2418 GTCACTTACGAAGGTACCATAGAGAAACCGCCAGGCATGCTACAGTGGATTTGCA 2477  
 Db 713 ValThrTyrGluGlyThrIleGluGlyAsnGlyArgGlyMetLeuGlnValAspPheAla 732  
 QY 2478 AACCGTTTCTGTCAGAGTGTGTAACAGTGCAGGACTGTGCAAGAAAGAAATCCGCTT 2537  
 Db 733 AsnArgPheValGlyGlyGlyValThrGlyAlaGlyLeuValGlnGluLysIleArgPhe 752  
 QY 2538 TTAATCAACCTGAGTGTGATTTTTCACGCTCTTCACTGAGTGTGATCAATGAA 2597  
 Db 753 LeuIleAsnProGluLeuIleValSerArgLeuPheThrGluValLeuAspHisAsnGlu 772  
 QY 2598 TGTCTTATCATCAGGTACTGACAGTACAGTGAATACAGAGCTATGCGGAAACATAC 2657  
 Db 773 CysLeuIleThrGlyThrGluGlnTyrSerGluTyrThrGlyTyrAlaGluThrTyr 792  
 QY 2658 CGCTGGCCCGGACCATGAGACAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2717  
 Db 793 ArgTyrAlaArgSerHisGluAspGlySerGluLysAspAspThrGlnArgArgCysThr 812  
 QY 2718 GAGATCGTCCCATCGACGCTCCCTCCATTCAGACGCTTACCTGACACCATGTTGTGCCGAG 2777  
 Db 813 GluIleValAlaIleAspAlaLeuHisPheArgTyrLeuAspGlnPheValProGlu 832  
 QY 2778 AAGATCAGCGGAGCTTAACAGGCTTACTGTGGATTTCTGCTGCTGAGTTCCTTCA 2837  
 Db 833 LysValArgArgGluLeuAsnLysAlaTyrCysGlyPheLeuArgProGlyValProSer 852  
 QY 2838 GAGAACCTGCTGCTGAGTGTGAGGAACTGGGGCTGTGTGCTGCTTGGGGGTGATGCT 2897  
 Db 853 GluAsnLeuSerAlaValAlaThrGlyAsnTyrGlyCysGlyAlaPheGlyGlyAspAla 872  
 QY 2898 AGACTAAAGCCTTAATACAGATCTCGCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2957  
 Db 873 ArgLeuLysAlaLeuIleGlnIleLeuAlaAlaAlaAlaGluArgAspValValTyr 892  
 QY 2958 TTCACTTTGGGAGCTCAGACTATGAGACATTTACAGCATGTCATACATCTCTCACT 3017  
 Db 893 PheThrPheGlyAspSerGluLeuMetArgAspIleTyrSerMetHisThrPheLeuThr 912  
 QY 3018 GAGAGGAACTGACTGTTGGAGAGTATATAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3077  
 Db 913 GluArgLysLeuAspValGlyLysValTyrLysLeuLeuLeuArgTyrTyrAsnGluGlu 932  
 QY 3078 TGCAGAACTGCTCCACCCCGGACAGACATCAAGCTTTTATCCATTCATACCATGCA 3137  
 Db 933 CysArgAsnCysSerThrProGlyProAspIleLysLeuTyrProPheIleTyrHisAla 952

QY 3138 GTTGAAGTCTCTGACACAGACACCAAC 3164  
 Db 953 ValGluSerSerAlaGluThrThrAsp 961  
 RESULT 7  
 Q9QYM2 PRELIMINARY; PRT; 972 AA.  
 AC Q9QYM2; 01-MAY-2000 (TREMREL. 13, Created)  
 DT 01-MAY-2000 (TREMREL. 13, Last sequence update)  
 DT 01-JUN-2003 (TREMREL. 24, Last annotation update)  
 DE Poly(ADP-ribose) glycohydrolase.  
 GN PARG.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BUF; TISSUE=Colon;  
 RA Shimokawa T., Masutani M., Nagasawa S., Nozaki T., Ikota N., Araki S.,  
 RA Aoki Y., Nakagawa H., Sugimura T.;  
 RT "Isolation and cloning of rat poly(ADP-ribose) glycohydrolase  
 (PARG)."  
 RL Submitted (Oct-1998) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AB019366; BAA87901.1; --  
 DR GO; GO:0016787; F:hydrolase activity; IEA.  
 DR InterPro; IPR007724; PARG.  
 DR Pfam; PF05028; PARG; 1.  
 KW Hydrolase.  
 SQ SEQUENCE 972 AA; 109002 MW; 3DED7885684F59E9 CRC64;  
 Alignment Scores:  
 Pred. No.: 0 Length: 972  
 Score: 4295.50 Matches: 808  
 Percent Similarity: 88.75% Conservative: 60  
 Best Local Similarity: 82.62% Mismatches: 101  
 Query Match: 59.31% Indels: 9  
 DB: 11 Gaps: 4  
 US-09-302-812-1 (1-4070) x Q9QYM2 (1-972)  
 QY 258 ATGAGTGGCGGCGCCGCTGTGAGCCCTGCACCAAGCAGCCCGCTGGGACGCGCTGCA 317  
 Db 1 MetSerAlaGlyProGlyCysGluProCysThrLysArgProArgTrpGlyAlaGly 20  
 QY 318 ACTTCTCCGCGCCGCTCGAGCCGCGAGTTCGCGGAGGAGGAGGAGGAGGAGGAGGAG 377  
 Db 21 ThrSerAlaProThrAlaSerAspSerArgSerPheProGlyArgGlnLysArgValLeu 40  
 QY 378 GATTCCAAAGGACGCTCGGTGACAGTTCAGGCTCCGCGCTCTCTGCTGAGCTGCGCCCTG 437  
 Db 41 AspProLysAspAlaProValGlnPheArgValProProSerSerSerAlaCysValSer 60  
 QY 438 GGCGCGCGGAGCAGCAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 497  
 Db 61 GlyArgAlaGlyProHisArgGlySerValThrSerPheValPheLysGlnLysProIle 80  
 QY 498 ACCAGTGTGATGACACTAAGGAATCAAGACAGTTCAGAAAGTTCAGAAAGTTCAGAAAG 557  
 Db 81 ThrThrTrpMetAspThrLysGlyProLysThrAlaGluSerGlu-----SerLys 97  
 QY 558 GAAACCAACATCAAGAGAGAAATCCATGATGAGTTCGTACAAAAGATACTTTTAT 617  
 Db 98 GluAsnAsnAsnThrArgThrAspProMetMetSerSerValGlnLysAspAsnPheTyr 117  
 QY 618 CAACATAACATGAAATAATAGAAATGTTCTCAGCTAGGTTTGTATAGTACCAGATT 677  
 Db 118 ProHisLysValGluLysLeuGlyAsnValProGlnLeuAsnLeuAspLysSerProThr 137  
 QY 678 GAAAGAGTACAGATATTTGAAGCAGCATCAGACTCGGCTATGCTAGTGGCAGAT 737  
 Db 138 GluLysSerThrProTyrLeuAsnGlnGlnThrAlaGlyValCysLysTrpHisSer 157



QY	738	GAAGGGCCACACTCAGAA	CGGCTTTTGGAAAGTGAACCTCCAGCGGTAACTCTGGTGACCA	797
Db	158	AlaGluValArgAlaGluSerAlaGluLeuSerAlaSerGluProSerAlaValThrGlnAlaPro	177	
QY	798	GAGCAGTTCAAGTAAATGCTAAATGTCATGCTGCTCCCAAAGATGATCACAAGTACACACA	857	
Db	178	LeuSerAsnAlaAsnAlaAsnGlnSerProThrAspGlyHisSerAspThr	197	
QY	858	AATAGTCAGAGAGTAGAGATAATCAGCAGTTTTCACACATGTAAAGCTCCGGAATGCA	917	
Db	198	AspHisGluGluAspArgAspAsnGlnGlnPheLeuThrProValIleuSerAlaAsnAla	217	
QY	918	AAGCAGACAGTGAAGATGAACAGGGCAGAGAAGCCAGAACCCACAGAAGTGTGGCAAG	977	
Db	218	LeuGlnThrValGlyAspGlyGln-----AlaArgSerAsnCysIysCysSerAla	234	
QY	978	GCTTGCATCTCTGCAGAAAGCCTGTGCGAGGTGTCTCAGCAGGAGAGACACAGCTGTGTCC	1037	
Db	235	SerCysGlnCysGlyGlnAspCysAlaGlyCysGlnArgGluGluAlaAspValIlePro	254	
QY	1038	GAGAGCCCTTGTCCGACACTGGCTCTCAGAGATGTTGTTACTCGACTGAAAATGCCAAC	1097	
Db	255	GluSerProLeuSerAspValGlyAlaGluAspIleGlyThrGlySerIleAsnAspAsn	274	
QY	1098	AGATTGAATAGACAAGAAAGTAGTCTAGGAAATTCCTCCATTTGAGAAGAAGTGA	1157	
Db	275	LeuLeuThrGlyGlnGluSerGlyLeuGlyAspSerProPheGluIleValGluSerGlu	294	
QY	1158	CCTGAGTCCACAAATGGATGTAGATAATCCAAAATAGTTGTCTCAGATTCCAGAAGCAGAT	1217	
Db	295	ProGluSerProMetAspValAspAsnSerIleThrSerCysGlnAspSerGluAlaAsp	314	
QY	1218	GAAGACACAAGTCCAGGTTTGTATGAACAGGAAGATAGCAGT-----TCTGCTCAACA	1271	
Db	315	GluGluAlaSerProValPheAspGluGlnAspAspGluAspArgSerSerGlnThr	334	
QY	1272	GCAATAAACCTTCAAGTTCACACCAAGAGAAGCTGACACTGAGTTGAGGAAGCGGTCC	1331	
Db	335	AlaAsnIleuSerArgGlnAlaArgGluValAspGlyAspLeuArgIleArgIleArgTyr	354	
QY	1332	TCTGCTAAGGAGGTGAGTTCGATTTCGATTTCCAATTTCAAGCAGGAGAGAGTCCAGCT	1391	
Db	355	LeuThrIleGlySerGluIleArgIleuHisPheGlnPheGluGlyGly---SerAsnAla	373	
QY	1392	GGAATGAATGATGTGAATGCCAAACGACCTCGAAGTACTTCTAGCTGAATGTAGAGTGC	1451	
Db	374	GlyThrSerAspLeuAsnAlaIleProSerGlyAsnSerSerSerLeuAsnValAspGly	393	
QY	1452	AGAAATTTCAAGCAATGGGGAAGAAGATTCTAAATCCAGATCATTTTCATGAGAGTG	1511	
Db	394	ArgSerSerIleGlnHisGlyIleArgAspSerIleIleThrAspHisPheValArgIle	413	
QY	1512	CCCAAAGCAGAGCAAAAGAAAGAAACAATGTGAATGAACATCAAGACACAGAAAGG	1571	
Db	414	ProIleSerGluAspIleArgIleGlyGlnCysGluValArgHisGlnArgAlaGluArg	433	
QY	1572	AGATCCCTTAATATACATTCACCTCACCTTTCTCCAGATAGAATGGCTTGGAACTCCT	1631	
Db	434	LysIleProIleTyrValProProAsnLeuProProAspIleIleTyrLeuGlyThrPro	453	
QY	1632	ATTGAGGAGATCAGAGAGATGCCAAGGTGTGGATCCGGCTCGCTCCCTTCCAGACCATCT	1691	
Db	454	IleGluGluMetArgIleMetProArgCysGlyValArgLeuProLeuLeuArgProSer	473	
QY	1692	GCAATATCACAGTGAATATTGGGTAGATCTTTGGCAATAGGGAAGTTCCTAAACCT	1751	
Db	474	AlaSerHisThrValThrValArgValAspLeuLeuArgAlaGlyGluValProIlePro	493	
QY	1752	TTCCCAACACATTTTAAAGATTTGTGGGACAAAGCATGTTAAGATCCCTTGTTCAGAA	1811	
Db	494	PheProThrHisTyrIleAspLeuIleProAspAsnIleHisValIleMetProCysSerGlu	513	

1812	QY	CAAACTCTGTACCTGTGGAAAGATGAGAAATGTGTGACGAGCTGACGAGCGCGGTGGGAA	1871
514	Db	GlnAsnLeuTyrProValGluAspGluAsnGlyGluArgThrAlaGlySerArgTrpGlu	533
1872	QY	CTCATTTACAGACTGCATCTCTCAACAGGCTCACTCGCGCCCAAGAACTCAAGATGCTATT	1931
534	Db	LeuileGlnThrAlaLeuLeuAsnLysPheThrArgProGlnAsnLeuLysAspAlaile	553
1932	QY	CTGAAGTACAATGTGGCATATTTCTAAGAAATGGGACTTTACAGCTTTGATTGATCTGG	1991
554	Db	LeuLysTyrAsnValAlaItyrSerLysLysTyrAspPheThrAlaLeuValAspPheTrp	573
1992	QY	GATAAGTACTAGAAGACGAGAGCTCAACACTGTGTATCACTCCATCTTGCTGCTGATATG	2051
574	Db	AspLysValLeuGluGluAlaGluAlaGlnHisLeuTyrGlnSerileuLeuProaspMet	593
2052	QY	GTGAAATTCACACTCTGTGCGCAATATTTGTACCCAGCCAAATACCACTCTCTGAAACAG	2111
594	Db	ValLysileAlaLeuCysLeuProAsnileCysThrGlnProileProleuLeuLysGln	613
2112	QY	AAGATCAATCATTTCCATCAATGTCCACAGGAACAGATTGCCAGTCTTTTACGTAATGCT	2171
614	Db	LysMetAsnHisSerValThrMetSerGlnGluGlnIleAlaSerLeuLeuAlaAsnAla	633
2172	QY	TTCTCTGCGAGTTTCCACGACCAATGCCAGATCAATCAAGTATTCAGATATTCAGTATCCA	2231
634	Db	PhePheCysThrPheProArgArgAsnAlaLysMetLysSerGluTyrSerSerIyPro	653
2232	QY	GATATTAACCTCAATCGTGTGTTTGAAGACGTTTCATCAAGGAAAAACACAGAAAGCTTAAA	2291
654	Db	AspIleAsnPheAsnArgLeuPheGluGlyArgSerSerArgLysProGluLysLeuLys	673
2292	QY	ACGCTCTCTGCTACTTTAGAAGAGTCACAGAGAAAAACCCACTGGTGTGTGACATTC	2351
674	Db	ThrLeuPheCysTyrPheArgArgValThrGluLysLysProThrGlyLeuValThrPhe	693
2352	QY	ACAAGA CAGAGTCTTGAAGATTTTCCAGAGTGGGAAGATGTCAAAAACCTCTGACTCGA	2411
694	Db	ThrArgGlnSerLeuGluAspPheProGluTrpGluArgCysAspLysProLeuThrArg	713
2412	QY	CTGCATCTCTACAGAGTACCATAGAAGGAAACGCGCAGGCGCATGTACAGSTGGAT	2471
714	Db	LeuHisValThrTyrGluGlyThrIleGluGlyAsnGlyArgGlyMetLeuGlnValAsp	733
2472	QY	TTTGCAAACCGTTTCGTTGGAGTGGTGTAAACAGTGCAGGACTGTGTGCAGAAGAAATC	2531
734	Db	PheAlaAsnArgPheValGlyGlyValThrGlyAlaGlyLeuValGlnGluIle	753
2532	QY	CGCTTTTAAATCAACCCCTGAGTGTATTGTTTCACGGCTCTTCACTGAGGCTCGGATCAC	2591
754	Db	ArgPheLeuIleAsnProGluLeuIleValSerArgLeuPheThrGluValLeuAspHis	773
2592	QY	AATGAATGCTTATCATCACAGGTACTGACGAGTACAGTACAATACACAGGCTATCGGAA	2651
774	Db	AsnGluCysLeuIleIleThrGlyThrGluGlnTyrSerGluTyrThrGlyTyrAlaGlu	793
2652	QY	ACATACCCCTCGGCGCCGAGCCATGAAGA CAGGAGCGAAAGGACGACTGGCAGAGCGC	2711
794	Db	ThrTyrArgTrpAlaArgSerHisGluAspGlySerGluLysAspAspTrpGlnArgCys	813
2712	QY	AGACTGATGATCGTCGCCATCGACGCCCTCCACTTCAGACGCTACCTCGACCACTTTGTG	2771
814	Db	CysThrGluIleValAlaIleAspAlaLeuHisPheArgArgTyrLeuAspGlnPheVal	833
2772	QY	CCCGAAGATGACAGCGGGAGCTTAAACAAGGCTTACTGTGGATTCTTCGCTCGGAGTT	2831
834	Db	ProGluLysValArgArgGluLeuAsnLysAlaItyrCysGlyPheLeuArgProGlyVal	853
2832	QY	TCTTCAGAGAAACCTGTCTGTCAGTGGCTACAGGAACTGGGGCTGTGCTGCTTGGGGGT	2891
854	Db	ProProGluAsnLeuSerAlaValAlaThrGlyAsnTrpGlyCysGlyAlaPheGlyGly	873
2892	QY	GATGCTAGACTAAAGCCCTTAATACAGATCTCTGGCAGCTGCTGTAGCTGACGAGACGTG	2951



```

874 AspAlaArgLeuYsAlaLeuLeuGlnLeuAlaAlaAlaAlaGluArgAspVal 893
2952 GTTATTTCACCTTTGGGAGCTCAGAACTGATGAGAGACATTACAGCATGATACATTC 3011
894 ValTyrPheThrPheGlyAspSerGluLeuMetArgAspIleTyrSerMetHisThrPhe 913
3012 CTCACCTGAGAGAACTGCTGTGAGAGAGTATATAAGCTGCTGCTACGATATTACAAAT 3071
914 LeuThrGluArgYsLeuAsnValGlyYsValTyrArgLeuLeuLeuArgTyrArg 933
3072 GAAGAATGACAGAACTGCTCCACCCCGGACAGACATCAAGCTTTATCCATTCAATAC 3131
934 GluGluCysArgAspCysSerProGlyProAspThrLysLeuTyrProPheIleTyr 953
3132 CATGAGTTGAGTCTCTGTACAGACCAACCCAGCGGACAAAGAGCGGG 3185
954 HisAlaAlaGluSerSerAlaGluThrSerAspGlnProGlyGlnArgThrGly 971

RESULT 8
Q8CB72 PRELIMINARY; PRT; 920 AA.
ID Q8CB72;
AC Q8CB72;
DT 01-MAR-2003 (T-EMBLrel. 23, Created)
DT 01-MAR-2003 (T-EMBLrel. 23, Last sequence update)
DT 01-JUN-2003 (T-EMBLrel. 24, Last annotation update)
DE Poly.
OS Mus musculus (Mouse).
GC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Bone;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium.
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RL 60,770 full-length cDNAs.";
RT Nature 420:563-573(2002).
DR ENBL; AK036656; BAC29519.1; -.
DR MGD; MGI:1347094; Parg.
DR InterPro; IPR007724; PARG.
DR Pfam; PF05028; PARG; 1.
SQ SEQUENCE 920 AA; 103799 MW; 379E79E121B09A1 CRC64;

Alignment Scores:
Pred. No.: 1.77e-311 Length: 920
Score: 4125.00 Matches: 782
Percent Similarity: 89.87% Conservative: 52
Best Local Similarity: 84.27% Mismatches: 86
Query Match: 56.96% Indels: 8
DB: 11 Gaps: 4

US-09-302-812-1 (1-4070) x Q8CB72 (1-920)
QY 258 ATGAGTGGGGCCCGCTGTGAGCCCTGACCAAGACCCCGCTGGAGCGCGCTGCA 317
DB 1 MetSerAlaGlyProGlyTyrGluProCysThrLysArgProArgTrpGlyAlaAlaGly 20
QY 318 ACTTCTCCGGCGCGCTCGGACCCCGGAGCTTCCCGGCGAGCAGAGCGCGTCTCTC 377
DB 21 ThrSerAlaProThrAlaSerAspSerArgSerPheProGlyArgGlnArgValLeu 40
QY 378 GATTCCAAAGAGCGTCCCGTGCAGTTCAGGTCGCGCTCTCTGCTCAGGTCGCGCTG 437
DB 41 AspProLysAspAlaProValGlnPheArgValProProSerSerProAlaCysValSer 60
QY 438 GCGCGGGGGACACACCGGAGCGCCACCTCTCTGTTTCAACAGAGAACTATA 497
DB 61 GlyArgAlaGlyProHisArgGlyAsnAlaThrSerPheValPheLysGlnLysThrIle 80

```

```

498 ACCAGTGGATGACACACTAAAGGATCAAGACAGTTGATCAGAAAGTTTGCATAGTAAA 557
81 ThrThrTrpMetAspThrLysGlyProLysThrAlaGluSerGlu-----SerLys 97
558 GAAACAAACAATCAAGAAGAATCCATGATGAGTCTCTGTACAAAGAAATACCTTTAT 617
98 GluAsnAsnAsnThrArgIleAspSerMetSerSerValGlnLysAspAsnPheTyr 117
618 CAACATAACATCGAAATAATTGAAAATGTTTCTCAGCTAGTGGTTTGTATAGTCACAGTT 677
118 ProHisLysValGluLysLeuGluAsnValProGlnLeuAsnLeuAspLysSerProThr 137
678 GAAAGAGGTGACACATATTGAGCAGCATCAGACTGCGGCTATGTGTAAAGTCGACAGAT 737
138 GluLysSerSerSerGlnTyrLeuAsnGlnGlnThrAlaSerValCysLysIleProGlnAsn 157
738 GAAGGGCCACACTCAGACAGCGCTTTTGGAAAGTGAACCTCCAGCGGTAACTCTGGTACCA 797
158 GluGlyLysHisAlaGluGlnLeuLeuAlaSerGluProProAlaGlyThrProLeuPro 177
798 GAGCAGTTCAGTAACTGCTTAATGTCATGATCGATCGTCCCCAAAGGATGATCACAAGTCACA 857
178 LysGlnLeuSerAsnAlaAsnIleGlyGlnSerProHisThrAspAspHisSerAspThr 197
858 AATAGTGAGGAGTAGAGATTAATCAGCAGTGTGTTTGACACATGTAAGCTTGCAGATGCA 917
198 AspHisGluGluAspArgAspAsnGlnGlnPheLeuThrProIleLysLeuAlaAsnThr 217
918 AAGCAGACGATGAGAACTGAACAGGGCAGAGAGCCAGAGCCAGCCAGAGTGTGCAAG 977
218 LysProThrValGlyAspGlyGln-----AlaArgSerAsnCysLysCysSerGly 234
978 GCTTGGCCATCCTGCAGAACCTGTGTGAGGTGTGACGAGGAGGAGACAGACGTGGTGTCC 1037
235 SerArgGlnSerValLysAspCysThrGlyCysGlnGlnGluValAspValLeuPro 254
1038 GAGAGCCCTTGTGCGACACTGGCTCTGAGGAGTGTGGTACTGAGCTGAGTGAATAATGCCAAC 1097
255 GluSerProLeuSerAspValGlyAlaGluAspIleGlyThrGlyProLysAsnAspAsn 274
1098 AGATTGAATAGACAAGAAAGTAGTCTTAGGAAATTCCTCCATTTGAGAAAGAAAGTGAA 1157
275 LysLeuThrGlyGlnGluSerSerSerLeuGlyAspSerProProPheGluLysGluSerGlu 294
1158 CTGAGTCCCAATGATGATGATGATTAATCCAAATAGTGTGTCAGGATTCAGAGATTCAGAG 1217
295 ProGluSerProMetAspValAspAsnSerLysAsnSerCysGlnAspSerGluAlaAsp 314
1218 GAAGAGACAAAGTCCAGGTTTGTGATGAACAGGAGATAGACAGTTCCTGCTCAACAGCAAT 1277
315 GluGluThrSerProValPheAspGluGlnAspAspArgSerSer---GlnThrAlaAsn 333
1278 AAACCTTCAAGGTTCCAAACCAAGAGAGCTGACACTGAGTTGAGGAGCGGTCTCTGCT 1337
334 LysLeuSerSerCysGlnAlaArgGluAlaAspGlyAspLeuArgLysArgTyrLeuThr 353
1338 AAGGAGGTGAGATTCGATTACATTTCCAAATTTGAAGGAGGAGAGTTCGAGTGGGAATG 1397
354 LysGlySerGluValArgLeuHisPheGlnPheGlu---GlyGluAsnAsnAlaGlyThr 372
1398 AATGATGTGAATGCAACACACCTGGAAGTACTTCTAGCCCTGAAATGTAGAGTGCAGAAAT 1457
373 SerAspLeuAsnAlaLysProSerGlyAsnSerSerSerLeuAsnValGluCysArgSer 392
1458 TCTAAGCAACATGGGAGAAAGGATTCTAAATTCAGATCATCATTTTCATGAGTGCAGTCCCAA 1517
393 SerLysGlnHisGlyLysArgAspSerLysIleThrAspHisPheMetArgIleSerLys 412
1518 GCAGAGACAAACAAAGAAACATGTGAATGAACATCAACAGCAACAGAAAGAGAGATC 1577
413 SerGluAspArgArgLysGluGlnCysGluValArgHisGlnArgThrGluArgLysIle 432
1578 CCTAAATACATTCCACCTCACCTTTCTCCAGATAAGAAATGGCTTGGAACTCTTATTGAG 1637

```







Db 452 AsnTrpGlyCysGlyAlaPheGlyAspSerTyrLeuLysAlaLeuLeuGlnLeuMet 471  
 QY 2925 GCAGCTGCTAGCTAGCAGCAGCTGTTTATTTTACCTTTGGGCACTCAGAACTGATG 2984  
 Db 472 ValCysAlaGlnLeuGlyArgProLeuAlaTyrTyrThrPheGlyAsnValGluPheArg 491  
 QY 2985 AGAGACATTTACAGATGATCATATCTCTCTAGAGAACTGCTGTTGGAGAGTA 3044  
 Db 492 AspAspPheHisGluMetTrpLeuLeuPheArgAsnAspGlyThrThrValGlnGlnLeu 511  
 QY 3045 TATAGCTGCTCTAGCTATTTACATGAGNATGAGAACTGCTCCACCCCGGACCA 3104  
 Db 512 TrpSer---IleLeuArgSerTyrSerArgLeuLeuGlySerSerLysGluPro 530  
 QY 3105 -----GACATCAAGCTTTTCCATTTCATATACATGCTGCTGAGTCC 3146  
 Db 531 ArgGluAsnLysAlaSerLysLysLysLeuTyrAspPheIle-----LysGluGlu 547  
 QY 3147 TGTACAGACACCAACACCGCGGACAAAGAGCGGGGCC 3188  
 Db 548 LeuLysLysValArgAspValProGlyGluGlyAlaSerAla 561  
 RESULT 11  
 Q9SKB3 PRELIMINARY; PRT; 548 AA.  
 AC Q9SKB3; Q94ET7;  
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Putative poly (ADP-ribose) glycohydrolase.  
 GN AT2G31870 OR TEJ.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsiis.  
 OC NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. Columbia;  
 RA Lin X., Kaul S., Shea T.P., Fujii C.Y., Shen M., VanAken S.E.,  
 RA Barnstead M.E., Mason T.M., Bowman C.L., Ronning C.M., Benito M.-I.,  
 RA Carrera A.J., Creasy T.H., Buell C.R., Town C.D., Nierman W.C.,  
 RA Fraser C.M., Venter J.C.;  
 RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Panda S., Poirier G.G., Kay S.A.;  
 RT "TEJ defines a role for poly-ADP-riboseylation in establishing period  
 RT length of the Arabidopsis circadian oscillator.";  
 RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AC006533; AAD32285.2; -;  
 DR EMBL; AF394690; AAK72256.1; -;  
 DR PIR; B84726; B84726.  
 DR GO; GO:0016787; F.hydrolase activity; IEA.  
 DR InterPro; IPR007724; PARG.  
 DR Pfam; PF05028; PARG; 1.  
 KW Hydrolase.  
 SQ SEQUENCE 548 AA; 62169 MW; 11A79FDA157C3329 CRC64;  
 Alignment Scores:  
 Pred. No.: 1.57e-36 Length: 548  
 Score: 585.00 Matches: 146  
 Percent Similarity: 46.78% Conservative: 65  
 Best Local Similarity: 32.37% Mismatches: 136  
 Query Match: 8.08% Indels: 104  
 DB: 10 Gaps: 10  
 US-09-302-812-1 (1-4070) x Q9SKB3 (1-548)  
 QY 1989 TGGGATAAGGTACTAGAGAGAGAGAGCTCAACACTTGTATCAGTCCATCTTGCCTGAT 2048  
 Db 79 PheAspGluLeuLeuAspGluLysGluSerLysArgTrpPheAspGluIleIleProAla 98

QY 2049 ATGGTGAATAATGACCTGTCTGCCAATATT----- 2081  
 Db 99 LeuAlaSerLeuLeuGlnPheProSerLeuLeuGluValHisPheGlnAsnAlaAsp 118  
 QY 2082 -----TGTACCAGCCATACCACTCTCTGAAACAGAGATGATCATTC 2126  
 Db 119 AsnIleValSerGlyIleLeuThrGlyLeuArgLeuLeuAsnSerGlnGlnAlaGlyIle 138  
 QY 2127 ATCAAAATGTCACAGAAACAGATTGCCAGTCTTTAGCTTAATGCTTTCTCTGCACGTTT 2186  
 Db 139 ValPheLeuSerGlnGluLeuIleGlyAlaLeuLeuAlaCysSerPhePheCysLeuPhe 158  
 QY 2187 CCACGACCAATGCCAAGATGAATCAGATTCAGATTCAGATTCAGATTCAGATTCAGATTC 2246  
 Db 159 ProAspAsnArgGlyAlaLys-----HisLeuProValIleAsnPheAsp 174  
 QY 2247 CGGTTGTTTGA-----GGACGTTTCATCAAGGAAACACAGAGAGCTTAAACG 2294  
 Db 175 HisLeuPheAlaSerLeuTyrIleSerTyrSerGlnSerGlnGlnSerLysIleArgCys 194  
 QY 2295 CTCCTCTCTACTTTAGAGAGTCACAGAGAAACCCACTGGGTTGTGTGACATTACA 2354  
 Db 195 IleMetHisTyrPheGluArgPheCysSerCysValProIleGlyIleValSerPheGlu 214  
 QY 2355 AGACAG-----AGTCTTGAAGATTTCCAGAGTGGGAAAGATGTGAAAAAATCTCTGACT 2408  
 Db 215 ArgLysIleThrAlaAlaProAspAlaAspPheTrpSerLysSerAspValSerLeuCys 234  
 QY 2409 CGACTGCTATCTCATTACCAAGGTACCATAGAGGAAACCGCCAGGCGATGTACAGGTG 2468  
 Db 235 AlaPheLysValHisSerPheGlyLeuIleGluAspGlnProAspAsnAlaLeuGluVal 254  
 QY 2469 GATTTGCAACCGTTTCTGGAGTGTGTAAACAGTCAGGACTTGTGCAAGAGAA 2528  
 Db 255 AspPheAlaAsnLysTyrIleuGlyGlySerLeuSerArgGlyCysValGlnGluGlu 274  
 QY 2529 ATCCGCTTTTAAATCAACCCCTGAGTTGTTTTCACGGCTCTTCACTGAGTGTCTGAT 2588  
 Db 275 IleArgPheMetIleAsnProGlnLeuIleAlaGlyMetLeuPheLeuProArgMetAsp 294  
 QY 2589 CACAATGATGCTTATCATCAGTACTGACAGTACTGACAGTGAATACAGAGCTATGCC 2648  
 Db 295 AspAsnGluAlaIleGluIleValGlyAlaGluArgPheSerCysTyrThrGlyTyrAla 314  
 QY 2649 GAAACATACCGTGGCGCGAGCCATGAGACAGAGCGAAGAGGACGACTGGCAGAGG 2708  
 Db 315 SerSerPheArgPheAlaGlyIleLysLysLysAlaMetAspProPheLysArg 334  
 QY 2709 CGCAGCTGAGTACTGCGCCATCGACGCCCTC-----CACTTCAGA 2750  
 Db 335 ArgArgThrArgIleValAlaIleAspAlaLeuCysThrProLysMetArgHisPheLys 354  
 QY 2751 CGCTACCTCGACCACTTTGTGCGGAGAGATCAGACGGGAGCTTAACAGGCTTACTGT 2810  
 Db 355 AspIleCys-----LeuLeuArgGluIleAsnLysAlaLeuCys 367  
 QY 2811 GGATTTCTTCGT----- 2822  
 Db 368 GlyPheLeuAsnCysSerLysAlaTrpGluHisGlnAsnIlePheMetAspGluGlyAsp 387  
 QY 2822 ----- 2822  
 Db 388 AsnGluIleGlnLeuValArgAsnGlyArgAspSerGlyLeuLeuArgThrGluThrThr 407  
 QY 2822 ----- 2822  
 Db 408 AlaSerHisArgThrProLeuAsnAspValGluMetAsnArgGluLysProAlaAsnAsn 427  
 QY 2823 -----CCTGGAGTTCTTCAGAGAACCTG-----TCTGCAGTG 2855  
 Db 428 LeuIleArgAspPheTyrValGluGlyValAspAsnGluAspHisGluAspAspGlyVal 447  
 QY 2856 GCTACAGGAACTGGGGCTGTGTGCTTGGGGGTGATGCTAGACTAAAGAGCTTAATA 2915

```

Db      448 AlaThrGlyAsnTrpGlyCysGlyValPheGlyGlyAspProGluLeuLysAlaThrIle 467
QY      2916 CAGATCCGCGAGCTGCTAGCTGAGCGAGCGAGTGGT---TATTTCACCTTTGGGAC 2972
Db      468 GlnTrpLeuAlaAlaSerGlnThrArgArgProPheIleSerTyrTyrThrPheGly--- 486
QY      2973 TCAGACTGATGAGACATTTACAGCATGCATACATTCCTCCTCAGTACGAGGAACTGACT 3032
Db      487 ValGluAlaLeuArgAsnLeuAspGlnValThrIleTrpIleLeuSerHisIleTrpThr 506
QY      3033 GTTGGAGAGTATATAAGCTGCTGCTACGATAT 3065
Db      507 ValGlyAspLeuTrpAsnMetMetLeuGluTyr 517

RESULT 12
Q8VYAL PRELIMINARY; PRT; 522 AA.
ID AC Q8VYAL
AC Q8VYAL
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Putative poly (ADP-ribose) glycohydrolase.
GN AT2G31870.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eucosids II; Brassicales; Brassicaceae; Arabidopsi.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Nguyen M., Karlin-Neumann G., Southwick A., Lam B., Miranda M.,
RA Palm C.J., Bowser L., Jones T., Banh J., Carninci P., Chen H.,
RA Cheuk R., Chung M.K., Hayashizaki Y., Ishida J., Kamiya A., Kawai J.,
RA Kim C., Lin J., Liu S.X., Naruoka M., Pham P.K., Sakano H.,
RA Sakurai T., Satou M., Seki M., Shinn P., Yamada K., Shinozaki K.,
RA Ecker J., Theologis A., Davis R.W.;
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY02330; AAL61937.1; -.
DR GO; GO:0016787; F:hydrolase activity; IEA.
DR InterPro; IPR007724; PARG.
DR Pfam; PF05028; PARG; 1.
KW Hydrolase.
SQ SEQUENCE 522 AA; 58942 MW; 6FF98685D867CA15 CRC64;

Alignment Scores:
Pred. No.: 1.7e-32 Length: 522
Score: 533.00 Matches: 132
Percent Similarity: 49.65% Conservative: 83
Best Local Similarity: 30.48% Mismatches: 138
Query Match: 7.36% Indels: 80
DB: 10 Gaps: 11

US-09-302-812-1 (1-4070) x Q8VYAL (1-522)
QY 1986 TTCTGGATAGGTACTAGACGAGCAGAGCTCAACACTTGTATCAGTCCATCTTGCT 2045
Db 80 PhePheAspLysIleSerArgGluGluSerAlaAsnPheGlyGluValPro 99
QY 2046 GATATGGTAAATTCACATCTGTCTGCCAAATATTGTACCAGCCATACCATCTCTG 2105
Db 100 AlaLeuCysArgLeuLeuGlnLeuProSerMetLeuGluLysHisTyr 116
QY 2106 AACACAGATGATCAT 2123
Db 117 ---GlnLysAlaAspHisValLeuAspGlyValLysSerGlyLeuArgLeuLeuGlyPro 135
QY 2124 -----TCCATCAATGTCCAGGAACAGATGCCAGTCTTTAGCTAATGCT 2171
Db 136 GlnGluAlaGlyIleValLeuLeuSerGlnGluLeuIleAlaLeuLeuAlaCysSer 155
QY 2172 TTCTTCTGCAGCTTCCACGACCAATGCCAAGATCAAGATCAAGATTCAGAGTATTCAGTTATCCA 2231

```

```

Db      156 PhePheCysLeuPheProGluValAspArgSerLeuLys-----AsnLeuGln 171
QY      2232 GATATTAACTTCAATCGTGTGTTTGAA-----GCAGCTTCATCAAGGAAACCA 2279
Db      172 GlyIleAsnPheSerGlyLeuPheSerPheProTyrMetArgHisCysThrLysGlnGlu 191
QY      2280 GAGAAGCTTAAACCGCTCTTCTGCTACTTAGAAGAGTCTACAGAGAAAACCCACTGGG 2339
Db      192 AsnLysIleLysCysLeuIleHisTyrPheGlyArgIleCysArgTrpMetProThrGly 211
QY      2340 TTGCTGACATTCACACAGACAGCTCTGAA-----GATTTT 2375
Db      212 PheValSerPheGluArgGlyIleLeuProLeuGluTyrHisProHisPheValSerTyr 231
QY      2376 CCAGAG-----TGGGAAAGATGTCAAAACCTCTGACTCGACTGCATGTCACTTAC 2426
Db      232 ProLysAlaAspSerTrpAlaAsnSerValThrProLeuCysSerIleGluIleHisThr 251
QY      2427 GAAGGTACCATAGAAGGAACGGCCAGGCATGTCTACAGGTGGATTTTGCAACCGTTTC 2486
Db      252 SerGlyAlaIleGluAspGlnProCysGluAlaLeuGluValAspPheAlaAspGluTyr 271
QY      2487 GTTGGAGTGTGTAAACCAGTGCAGGACTTCTGCAAGAGAAATCCGCTTTTAAATCAAC 2546
Db      272 PheGlyGlyLeuThrLeuSerTyrAspThrLeuGlnGluIleArgPheValIleAsn 291
QY      2547 CCTGAGTTGATTGTTTTCACGGCTCTTCACTGAGGTGCTGGATCAACATGAATGCTTATC 2606
Db      292 ProGluLeuIleAlaGlyMetIlePheLeuProArgMetAspAlaAsnGluAlaIleGlu 311
QY      2607 ATCAGAGTACTGAGCAGTACAGTGAATACACAGGCTATGCGGAACATACCGTGGGCC 2666
Db      312 IleValGlyValGluArgPheSerGlyTyrThrGlyTyrGlyProSerPheGlnTyrAla 331
QY      2667 CGAGCCCATGAGCAGGAGCGAAGGACGACTGGCAGAGCGCAGCAGCTGAGATCGTC 2726
Db      332 GlyAspTyrThrAspAsnLysAspLeuAspIlePheArgArgGlyThrArgValIle 351
QY      2727 GCCATCGAGCGCTCCACTTCAGACGCTACCTCCAGCAGTGTTCGCCGAGAGATCAGA 2786
Db      352 AlaIleAspAlaMetPro---AspProGlyMetGlyGlnTyrLysLeuAspAlaLeuIle 370
QY      2787 CGGAGCTTAAACAGGCTTACTGTGGATTCTTCGT----- 2822
Db      371 ArgGluValAsnLysAlaPheSerGlyTyrMetHisGlnCysLysTyrAsnIleAspVal 390
QY      2823 -----CTGGAGTTTCTTCA----- 2837
Db      391 LysHisAspProGluAlaSerSerSerHisValProLeuThrSerAspSerAlaSerGln 410
QY      2838 -----GAGAACCTGTCTCCAGTGGCT 2858
Db      411 ValIleGluSerSerHisArgTrpCysIleAspHisGluGluLysLysIleGlyValAla 430
QY      2859 ACAGAAAACGCGCTGTGTGTCCTTTGCGGTGTATGTAGACTAAAGCCCTTAATACAG 2918
Db      431 ThrGlyAsnTrpGlyCysGlyValPheGlyGlyAspProGluLeuLysIleMetLeuGln 450
QY      2919 ATCTCTGGCAGCTGTGTAGTGTGAGCAGCAGCTGGT---TATTTCACCTTTGGGACTCA 2975
Db      451 TrpLeuAlaIleSerGlnSerGlyArgProPheMetSerTyrTyrThrPheGly---Leu 469
QY      2976 GAACTGAGAGACATTTACAGCATGCATACATTCCTCAGTACGAGAGAAACCTGCTGT 3035
Db      470 GlnAlaLeuGlnAsnLeuAsnGlnValIleGluMetValAlaLeuGlnGluMetThrVal 489
QY      3036 GAAGAAGTATATAACTGCTGCTGCTACGATATTAACAATGAA 3074
Db      490 GlyAspLeuTrpLysLysLeuValGluTyrSerSerGlu 502

RESULT 13
Q867X0 PRELIMINARY; PRT; 781 AA.
ID Q867X0

```

AC Q867X0;  
 DT 01-JUN-2003 (TrEMBLrel. 24, Created)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Poly ADP-ribose metabolism enzyme-3 long form (C. elegans PME-3  
 DE protein) (Corresponding sequence F20C5.1a).  
 GN F20C5.1 OR PME-3.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;  
 OC Rhabditidae; Peloderinae; Caenorhabditis.  
 OX NCBI\_TaxID=6239;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Gagnon S.N., Hardy I., Desnoyers S.;  
 RT "Characterization of poly(ADP-ribose) glycohydrolases in the nematode  
 RT Caenorhabditis elegans";  
 RL Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=99069613; PubMed=9851916;  
 RA none;  
 RT "Genome sequence of the nematode C.elegans: A platform for  
 RT investigating biology";  
 RL Science 282:2012-2018 (1998).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RA Matthews P.;  
 RL Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AV185493; AAC26316.1; --  
 DR EMBL; Z68161; CAD89735.1; --  
 DR WormPep; F20C5.1a; CE33775.  
 DR InterPro; IPR007724; PARG.  
 DR Pfam; PF05028; PARG; 1.  
 SQ SEQUENCE 781 AA; 89273 MW; 30447D06605CED0A CRC64;

Alignment Scores:  
 Pred. No.: 1.95e-28 Length: 781  
 Score: 481.50 Matches: 210  
 Percent Similarity: 38.53% Conservative: 126  
 Best Local Similarity: 24.08% Mismatches: 297  
 Query Match: 6.65% Indels: 239  
 DB: Gaps: 37

US-09-302-812-1 (1-4070) x Q867X0 (1-781)

QY 852 GACACAAATAGTGAGGAGTAGAGATAATACAGCAGTTTTTGACATGTAAGCTTGGC 911  
 DB 10 AspProValThrGlnAspGluLysAspTyrGluAspTyrVal---GlyValGlyPheAla 28  
 QY 912 AATGCAAGCAGCAGATGGAA-----GATGAACAGGCAGCAGAGCCAGAGCC 959  
 DB 29 HisGlnValProThrMetLysArgArgLysLeuThrGluHisGlyAsnThrThrGluSer 48  
 QY 960 CACAGAAAGTGTGCAAGGCTTGCCATCTCGCAGAGCCCTGTGCAGGCTGTACAGCAGGAG 1019  
 DB 49 LysGlu-----AspProGluGlu-----ProLys 56  
 QY 1020 GAGACACAGACCTGTGTCCGAGAGCCCTGTGCGACACTGGCTCTGAGGAT----- 1070  
 DB 57 SerArgAspValPheValSerSerGlnSerSerAspGluSerGlnGluAspSerAlaGlu 76  
 QY 1071 -----GTTGGTACTGGACTGAAATGCC----- 1094  
 DB 77 AsnProGluLeuAlaLysGluValSerGluAsnCysGluAsnLeuThrGluThrLeuLys 96  
 QY 1095 -----AACAGATTGAATAGACAGAAAGTAGCTAGTAGAAT 1130  
 DB 97 IleSerAsnIleGluSerLeuAspAsnValThrGluArgSerGluHisThrLeuAspAsn 116  
 QY 1131 -----TCTCCTCCATTGAGAAAGAAAGTAGTGAACCTGAGTCCCAATGGATGTAGAT 1181  
 DB 117 HisLysSerThrGluProMetGluGluAspValAsnAsnLysSerAsnIleAspVal--- 135

QY 1182 AATCCAAAATAGTTGTCCAGGATTCAAGACGA-----CATGAAGAGACA 1226  
 DB 136 ---AlaIleAsnSerAspGluAspAspGluLeuValLeuGluAsnAsnLysGluMet 154  
 QY 1227 AGTCAGGTTTTGATGAACAGGAGAT----- 1253  
 DB 155 ArgAspGlyGluGlnValGlnGlnAspLeuPheAlaAspGlnGluLeuIleGluTyr 174  
 QY 1254 -----AGCAGTTCTGCTCAACA 1271  
 DB 175 ProGlyIleMetLysAspThrThrGlnLeuAspIleThrAspSerGluValGluThr 194  
 QY 1272 GCAATATAACCTTCAAGGTTCCAAACCAAGAGAAGCTGACACTGAGTTGAGGAAGCGGTCC 1331  
 DB 195 AlaGlnLysMetGluMetIleGluGluThrGluAlaAspSer----- 208  
 QY 1332 TCTGCTAAGGAGGTGAGATTCGATTACATTTTCAATTTGAAGGAGGAGAGTCCAGCT 1391  
 DB 209 ---ThrPheValGlyGluAspSerLys--- 216  
 QY 1392 GGAATGAATGATGTGAATGCCAAACGACCTGGAAGTACT----- 1430  
 DB 217 -----AsnGlnArgGlnSerGlyThrThrSerAspGluValAspAlaAsp 231  
 QY 1431 TCTAGCCTGAATGTAGAGTGCAGAAATTTCTAAGCAACATGGGAGAAAGGATTCTAAAAATC 1490  
 DB 232 SerGlnIleAsnLeuAlaThrLysThrValArg-----ThrSerSer 245  
 QY 1491 ACAGATCATTTGATGAGATGCCCAAGCAGAGGACAAAGAAAGAAACAATGTGAATG 1550  
 DB 246 SerSerPheLeuSerThrValSerThrCysGluAlaProAlaLysGlyArgAlaArgMet 265  
 QY 1551 AAACATCAAGAACAAGAGGAGAGTCCCTAAATACATATCCACCTCACCTTTCT----- 1604  
 DB 266 TyrGlnLysGluLeuGluLysHisValIleAlaPheThrGluGlyAsnLeuThrLeuGln 285  
 QY 1605 CCAGATAAGAAATGGCTTGGAACTCTTATTGAGGAGATGAGGAGAAATGCCAAGTGTGGG 1664  
 DB 286 ProAsp----- 287  
 QY 1665 ATCCGGCTCCCTCCCTTGGAGACCATCTGCCAATCACAACAGTACTATTCGGGTAGATCTT 1724  
 DB 288 ---LeuAsnLysValAspProAspArgAsnTyr----- 297  
 QY 1725 TTGCGAATAGGAGAGTCTCTAAACCTTCCCA-----ACACATTTTAAAGATTGTGGGAC 1781  
 DB 298 ---ArgTyrCysThrIleProAsn---PheProAlaSerGlnGlyLysLeuArgGluAsp 315  
 QY 1782 AACACAGCAT-----GTTAAGATCCCTTGTTCAGAACAAACTTGTACCCCTGTGGAA 1832  
 DB 316 AsnArgTyrGlyProLysIleValLeuProGlnArgTipArgLysPhe----- 331  
 QY 1833 GATGAGAAATGGTGGAGAGCTGCGAGCAGCGGCGGAGGAACTCATTGAGTGCACCTCTC 1892  
 DB 332 AspSerArgGlyArg----- 337  
 QY 1893 AACAGGCTCACTCGGCCCCAGACCTGAAGATGCTATTCTGAAGTACAAT----- 1943  
 DB 338 ---ArgAspSerTyrPheTyrPheLysArgLysLeu 348  
 QY 1944 GTGGCATATCTTAAGAAATGGGACTTTTACAGCT-----TTGATT 1982  
 DB 349 AspGlyTyrLeuLysCysTyrLysThrThrGlyTyrPheMetPheValGlyLeuLeuHis 368  
 QY 1983 GATTTCGGAT-----AAGGTACTGAGAGAGCAGAGCTCAACACTTGTATCAG 2033  
 DB 369 AsnMetTyrGluPheAspProAspIleThrTyrLysLeuProAlaLeuGluMetTyrTyr 388  
 QY 2034 TCCATCTGCTCATATGGTG----- 2054  
 DB 389 LysGluMetSerGluLeuValGlyArgGluGluValLeuLeuLysPheAlaArgValAla 408  
 QY 2055 AAAATTGCACTCTGTCTGCCAAATATTGTTACCCAGCCCAATACCCTCTCTGAAACAGAAG 2114



Db 745 GlyLeuProHisLysHisPheGluPheTyrValPhe 756

RESULT 14

Q19637 PRELIMINARY; PRT; 764 AA.

AC Q19637; SEQUENCE FROM N.A.

DT 01-NOV-1996 (TrEMBLrel. 01, Created)

DT 01-MAR-2003 (TrEMBLrel. 23, last sequence update)

DT 01-OCT-2003 (TrEMBLrel. 25, last annotation update)

DE PME-3 protein (corresponding sequence F20C5.1) (Poly ADP-ribose

DE metabolism enzyme-3 short form).

DE F20C5.1 OR PME-3.

GN Caenorhabditis elegans.

OS Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;

OC Rhabditidae; Peloderinae; Caenorhabditis.

OC NCBI\_TaxID=6239;

RN [1]

RN SEQUENCE FROM N.A.

RP Matthews P.;

RA Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.

RL [2]

RL SEQUENCE FROM N.A.

RE MEDLINE=99069613; PubMed=9851916;

EX none;

RT "Genome sequence of the nematode C.elegans: A platform for

RT investigating biology.;"

RT Science 282:2012-2018 (1998).

RN [3]

RN SEQUENCE FROM N.A.

RA Gagnon S.N., Hardy I., Deaneviers S.;

RT "Characterization of poly (ADP-ribose) glycohydrolases in the nematode

RT Caenorhabditis elegans.;"

RL Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.

DR EMBL; Z68161; CAA92299.2; -.

DR EMBL; AY185494; AAC026317.1; -.

DR PIR; T21138; T21138.

DR WormPep; F20C5.1b; CE32867.

DR InterPro; IPR007724; PARG.

DR Pfam; PF05028; PARG; 1.

SQ SEQUENCE 764 AA; 87441 MW; B391703B1D0CB79A CRC64;

Alignment Scores:

Pred. No.: 6.19e-28 Length: 764

Score: 475.00 Matches: 208

Percent Similarity: 37.88% Conservative: 120

Best Local Similarity: 24.02% Mismatches: 294

Query Match: 6.56% Indels: 244

DB: Gaps: 36

US-09-302-812-1 (1-4070) x Q19637 (1-764)

QY 852 GACACAATAGTGGAGGAGTAGAGATAATCAGCAGTTTTTGACACATGTAAGCTTGCG 911

Db 10 AspProValThrGlnAspGluLysAspTyrGluAspTyrVal---GlyValGlyPheAla 28

QY 912 AATGCAAGCAGCAGCATGTGAA-----GATGAACAGGGCAGCAGCAGCAGAGC 959

Db 29 HisGlnValProThrMetLysArgGlyLysLeuThrGluHisGlyAsnThrThrGluSer 48

QY 960 CACACAGAGTGTGGCAGGCTTGCCATCTCCAGACGCTGTGCAGGCTGTGCAGGAGG 1019

Db 49 LysGlu-----AspProGluGlu-----Prols 56

QY 1020 GAGACAGACGCTGGTGTCCAGAGCCCTTCTCGGACACTGGCTCTGAGGAT----- 1070

Db 57 SerArgAspValPheValSerSerGlnSerSerAspGluSerGlnGluAspSerAlaGlu 76

QY 1071 -----GTTGGTACTGGAGCTGAAATGCC----- 1094

Db 77 AsnProGluIleAlaLysGluValSerGluAsnCysGluAsnLeuThrGluThrLeuLys 96

QY 1095 -----AACAGATTGAATAGACAAGAAAGTAGTCTAGGAAT 1130



Db 97 IleSerAsnIleGluSerLeuAspValThrGluArgSerGluHisThrLeuAspAsn 116  
QY 1131 -----TCCTCTCCATTTTTCAGAAAGAAAGTGAACCTGAGTCAACAAATGATGTA--- 1178  
Db 117 HisLysSerThrGluProMetGluGluAspValAsnAsnLysSerAsnIleAspValAla 136  
QY 1179 -----GATAATTCGAAATAGTTGTCAGGAT 1205  
Db 137 IleAsnSerAspGluAspAspGluLeuValLeuGluGluAsnLysGluMetArgAsp 156  
QY 1206 TCAGAGCAGATGAAGACACAGTCCAGGTTT-----GATGAACAGGAA--- 1250  
Db 157 GlyGluGlnValGlnGlnSerGlnAspLeuPheAlaAspGlnGluLeuGlu 176  
QY 1251 -----GATAGCAGATTCTGCTCAA 1268  
Db 177 TyrProGlyIleMetLysAspThrThrThrGlnLeuAspIleThrAspSerGluValGlu 196  
QY 1269 ACAGCAAAATAAACCCTTCAAGTTCCAAACCAAGAGAGCTGACACTGAGTTGAGGAGCGG 1328  
Db 197 ThrAlaGlnLysMetGluMetIleGluGluThrGluAlaAspSer----- 211  
QY 1329 TCCTCTGCTAAGGAGGTGAGATTGATTACATTTCCAAATTTGAAGGAGGAGAGTCCA 1388  
Db 212 -----ThrPheValGlyGluAspSerLys 219  
QY 1389 GCTGGAATGAATGATGTAATGCCAAACGACCTGGAAGTACTTCTAGCCCTGAATGAGAG 1448  
Db 220 Ala-----ThrLysThrValArgThrSerSerSerPheLeuSerThr--- 234  
QY 1449 TCCAGAAATCTAAGCAACATCGGAGAAAGGATCTAAATACACAGATCATTTTCATGAGA 1508  
Db 234 ----- 234  
QY 1509 GTGCCCAAGCAGACAGCAAAAGAAAGAACATGTAATGAACATCAAGAAACAGAA 1568  
Db 235 ValSerThrCysGluAlaProAlaLysGlyArgAlaArgMetTyrGlnLysGluLeuGlu 254  
QY 1569 AGGAAGATCCCTAAATACATTCACCTCCACCTTTCT-----CCAGATAAGAAATGGCTT 1622  
Db 255 LysHisValIleAlaPheThrGluGlyAsnLeuThrLeuGlnProAsp----- 270  
QY 1623 GGAATCCTATTGAGGAGATGAGGAGAAATGCCNAGGTGGGATCCGGTGCCTCCCTG 1682  
Db 271 -----LeuAsnLysVal 274  
QY 1683 AGACCATCTGCCAATCACACAGTGAATCTTCGGGTAGATCTTTTGGCAATAGGAGAGTT 1742  
Db 275 AspProAspArgAsnTyr-----ArgTyrCysThrIle 285  
QY 1743 CCTAAACCTTTCCCA---ACACATTTTAAAGATTTGTGGGACACAAAGCAT----- 1790  
Db 286 ProAsn---PheProAlaSerGlnGlyLysLeuArgLysAspAsnArgTyrGlyProLys 304  
QY 1791 GTTAAGATCGCTTGTTCAGAAACAACTGTACCTCTGTGGAAGATCAGATGTTGAGCGA 1850  
Db 305 IleValLeuProGlnArgTyrArgGluPhe-----AspSerArgGlyArg 320  
QY 1851 GCTGCAGGCGCGGTGGGAACCTCATTCAGACTGCATCTTCAACAGGCTCACTCGGCC 1910  
Db 320 ----- 320  
QY 1911 CAGAACCTGAAGATGCTATTCTGAAGTACAAT-----GTGGCATATTTCAAGAAA 1961  
Db 321 -----ArgAspSerTyrPheTyrPheLysArgLysLeuAspGlyTyrLeuLysCys 337  
QY 1962 TGGACCTTTACAGCT-----TTGATTGATTTCTGGAT----- 1994  
Db 338 TyrLysThrThrGlyTyrPheMetPheValGlyLeuLeuHisAsnMetTyrGluPheAsp 357  
QY 1995 ---AAGTACTAGAAGAGCAGAGCTCAACACTTGTATCAGTCCATCTTTCCTGATGATG 2051  
Db 358 ProAspIleThrTyrLysLeuProAlaLeuGluMetTyrTyrLysGluMetSerGluLeu 377

QY 2052 GTG-----AAAATTGCACCTCTGCTCG 2072  
Db 378 ValGlyArgGluGluValLeuGluLysPheAlaArgValAlaAlaLysThrAla 397  
QY 2073 CCAATATTTTGTACCAGCAATACCACTCTCTGAAACAGAGATGAATCATTTCCATCACA 2132  
Db 398 GluAspIleLeuProGluArgIleTyrArgLeuValGlyAspVal---GluSerAlaThr 416  
QY 2133 ATGTCCAGGAAACAGATTGCCAGCTTTTAGCTAATAGTCTTTCTTCGACGTTTCCACGA 2192  
Db 417 LeuSerHisLysGlnCysAlaAlaLeuValAlaArgMetPhePhe----- 431  
QY 2193 CGCAATGCCAGATGAATACAGATATTCCAGTTATCCAGATTATTAACCTCAATCGGTG 2252  
Db 432 -----AlaArgProAspSerPheSer-----PheCysArgIle 443  
QY 2253 TTTGAAGGAGCTTATCATCAAGGAAACAGAGAACTTAAACCGCTCTTCTGTACTTTAGA 2312  
Db 444 LeuSerSerAspLysSerIleCysValGluLysLeuLysPheLeuPheThrTyrPheAsp 463  
QY 2313 AGAGTCACAGAGAAAAAACCCCACTGGGTGTGCATTCTC-----ACAAGACAGAGCTCT 2366  
Db 464 LysMetSerMetAspProProAspGlyAlaValSerPheArgLeuThrLysMetAspLys 483  
QY 2367 GAAGATTTTCAGAGTGGGAAAGATGTGAATACTCCTGACTCGACTGCTCATCTTAC 2426  
Db 484 AspThrPheAsnGluGluTyrLysAspLysLysLeuArgSerLeuProGluValGluPhe 503  
QY 2427 -----GAAGGTACCATAGAGAAACCGCCAGGCGATCTACAGGTGCTGATCACAATGAATCT 2600  
Db 504 PheAspGluMetLeuIleGluAspThrAlaLeu---CysThrGlnValAspPheAlaAsn 522  
QY 2481 CTTTTCGTGGAGTGGTGAACACAGTCGAGACTTGTGCAAGAGAAATCCGCTTTTAA 2540  
Db 523 GluHisLeuGlyGlyValLeuAsnHisGlySerValGlnGluIleArgPheLeu 542  
QY 2541 ATCAACCTCGATTGATTGTTTTCACGGCTCTTCACGTGAGTCTGATCACAATGAATCT 2600  
Db 543 MetCysProGluMetMetValGlyMetLeuLysGlyLysMetLysGlnLeuGluAla 562  
QY 2601 CTTATCATCACAGGTACTGAGCAGTACAGTGAATACACAGGTCTATCCGAAACATACC 2660  
Db 563 IleSerIleValGlyAlaTyrValPheSerSerTyrThrGlyTyrGlyHisThrLeuLys 582  
QY 2661 TGGGCT-----CGAGCCATGAAGACAGAGAGCGAA---AGGACAGACTCG 2702  
Db 583 TrpAlaGluLeuGlnProAsnHisSerArgGlnAsnThrAsnGluPheArgAspArgPhe 602  
QY 2703 CAGAGCGCACAGCTGAGATCGTCCCATCGACCTCCACTTTCAGACGC----- 2753  
Db 603 GlyArgLeuArgValGluThrIleAlaIleAspAlaIleLeuPheLysLysLeu 622  
QY 2754 -----TACCTCGACAGTTTGTGCCGAGAAAGATCAGACGGGAGCTTAAACAGGCTTAC 2807  
Db 623 AspCysGlnThrGluGlnLeuAsnLysAlaAsnIleIleArgGluMetLysLysAlaSer 642  
QY 2808 TGTGAATTTCTCGTCTCGAGTCTTCTTCAGAGAACCTCTTCGAGTGGCTACAGAAAC 2867  
Db 643 IleGlyPheMetSerGlnGlyProLysPheThrAsnIle---ProIleValThrGlyTyr 661  
QY 2868 TGGGCTGTGGCTCTTGGGGTGTAGTACTTAAAGCCCTTAAATACAGATCTCTGGCA 2927  
Db 662 TrpGlyCysGlyAlaPheAsnGlyAspLysProLeuLysPheIleIleGlnValIleAla 681  
QY 2928 GTGTGTAGCTGAGCAGACGTGGTTTATTTCACCTTTGGGAGCTCAGAACTGATGAGA 2987  
Db 682 AlaGlyValAlaAspArgProLeuHisPheCysSerPheGlyGluProGluLeuAla 701  
QY 2988 GACATTTACAGATGATCATCTCTCCTGAGAGAACTGACTGTTGGAGAGTATAT 3047  
Db 702 LysCysLysIleIleGluArgMetLysGlnLysAspValThrLeuGlyMetLeuPhe 721

QY 3048 AAGCTGCTGCTACGATATTACATGAAGATGCAGAAACTGCTCCACCCCGGACACGAC 3107  
 Db 722 SerMetile-----AsnSerThrGlyLeuProHisLysHis 733

QY 3108 ATCAAGCTTTATCCATTC 3125  
 Db 734 PheGluPheTyValPhe 739

RESULT 15  
 Q9NSL4 PRELIMINARY; PRT; 485 AA.  
 AC Q9NSL4;  
 DT 01-OCT-2000 (TRENBLrel. 15, Created)  
 DT 01-MAR-2003 (TRENBLrel. 23, Last sequence update)  
 DT 01-JUN-2003 (TRENBLrel. 24, Last annotation update)  
 DE Hypothetical protein H23L24.5 (Poly ADP-ribose metabolism enzyme 4).  
 DE H23L24.5 OR PME-4.  
 GN Caenorhabditis elegans.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
 OC Rhabditidae; Peloderinae; Caenorhabditis.  
 OX NCBI\_taxid=6239;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Bristol N2;  
 RX MEDLINE=99069613; PubMed=9851916;  
 RA Waterston R.;  
 RT "Genome sequence of the nematode C. elegans: a platform for  
 RT investigating biology. The C. elegans Sequencing Consortium.";  
 RL Science 282:2012-2018(1998).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Bristol N2;  
 RX Abu-Threideh J.; Lehnert L.;  
 RA "The sequence of C. elegans cosmid H23L24.";  
 RT Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Bristol N2;  
 RA Waterston R.;  
 RL Submitted (DEC-2002) to the EMBL/GenBank/DBJ databases.  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RA Gagnon S.N., Hardy I., Desnoyers S.;  
 RT "Molecular characterization of poly ADP-ribose glycohydrolases in  
 RT Caenorhabditis elegans.";  
 RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AC006662; AAF39896.2; -;  
 DR EMBL; AF548468; AAN40699.1; -;  
 DR WormPep; H23L24.5; CE32685.  
 DR InterPro; IPR007724; PARG.  
 DR Pfam; PF05028; PARG; 1.  
 KW Hypothetical protein.  
 SQ SEQUENCE 485 AA; 55574 MW; 86367DDBAB15081F CRC64;

Alignment Scores:  
 Pred. No.: 9,08e-28 Length: 485  
 Score: 472.00 Matches: 143  
 Percent Similarity: 47.18% Conservative: 83  
 Best Local Similarity: 29.85% Mismatches: 165  
 Query Match: 6.52% Indels: 88  
 DB: 5 Gaps: 17

US-09-302-812-1 (1-4070) x Q9NSL4 (1-485)

QY 1866 TGGGAACCTCAGCTGCTCTCAACAGCTCACTCGG---CCCCAGAACCTGAG 1922  
 Db 46 PheGluLeuGluThrGlyValSerGlnGlnTrpArgAsnCysAspGlnAsnLeuPhe 65  
 QY 1923 GATGCTATTCTGAAG-----TACAACTGGGCATATTCT----- 1955  
 Db 66 AsnGluTyLeuLysThrTyLysAsnGlyTyLysSerGlnPheGluAspLeuLeuPhe 85

QY 1956 -----AAGAAATGGGACTTTTACAGTTTGTGATTGATTTC 1988  
 Db 86 LysIleTrpGlyTySerGluGluGlyAspLeuProAlaLeuLysSerPhe 105  
 QY 1989 TGGGATAGTACTAGNA-----GAAGCAGAGCTCAACACTTGTATCAGTCCATC 2039  
 Db 106 TyArgLysMetSerGluIleValGlyGluAspGlu-----Val 118  
 QY 2040 TTGCCTGATATGGTGAATAATTGCACCTCTGTCTGCCAAATATTTGTACCCAGCCAATACCA 2099  
 Db 119 LeuGluLysLeuAlaArgLeuValArgIleThrLysSerAlaCys-----Glu 134  
 QY 2100 CTCTGAAACACAGAGATGAAT-----CATTCATCACAATGTCACAG 2141  
 Db 135 ValLeuProGluLysIleTyArgLeuValGlyAspIleGluSerAlaThrPheSerHis 154  
 QY 2142 GAACAGATTGCGAGTCTTTTAGCTTAATGCTTCTTCCGCGTTCACACGCAATGCC 2201  
 Db 155 IleGlnCysAlaSerLeuIleAlaTrpMetPhePheSerAspThrProArg----- 171  
 QY 2202 AAGATGAATCAGAGTATTCCAGTTATCCAGATATTAACTTCAATCGTTGTTTGAAGGA 2261  
 Db 172 -----LeuSerPheIleIleLeuGlnLys 180  
 QY 2262 GGTTCATCAAGAAACACAGAGACTTAAACGGCTCTTCTGCTACTTTAGAGAGTCA 2321  
 Db 181 ThrThrCysValAlaValGluLysLeuLysPheLeuPheThrTyPheAspLysMetSer 200  
 QY 2322 GAGAAACCCACTGGTGGTGCACATTCA-----AGCAGAGT 2363  
 Db 201 IleAspProProlIleGlyAlaValSerPheArgLysMetArgIleThrHisLysGlnTy 220  
 QY 2364 CTTCAGAGATTTCCAGAGTGGGAA---AGATGTGAAACACTCCGACTCGACTCGATGTC 2420  
 Db 221 LeuGluAsn-----TrpLysLeuArgGluThrAsnLeuLeuProAspValGlnVal 237  
 QY 2421 ACTTACGAGGTACCATAGAGGAAACGGCCAGGCATGCTACAGTGTGATTGTTGCAAC 2480  
 Db 238 PheAspLysMetSerIleGluGluThrAlaLeu---CysThrGlnIleAspPheAlaAsn 256  
 QY 2481 CGTTTCGTTGGAGCTGTGTAAACAGCTGCAGACTGTGTGCAAGAGAAATCCGCTTTTA 2540  
 Db 257 LysArgLeuGlyGlyGlyValLeuLysGlyAlaValGlnGluGluIleArgPheMet 276  
 QY 2541 ATCAACCTGAGTTGATTGTTTTCACGGCTCTTCTACTAGGTGCTGGATCACAATGAATGT 2600  
 Db 277 MetCysProGluMetMetValAlaIleLeuAsnAspValThrGlnAspLeuGluAla 296  
 QY 2601 CTTATCATCACAGGTACTGACGACATGACAGTGAATACACAGGCTATGCCGAAACATACCGC 2660  
 Db 297 IleSerIleValGlyAlaTyValPheSerSerTyThrGlyTySerAsnThrLysLys 316  
 QY 2661 TGGGCCCGG-----AGCCATGAAGACAGGACGCGAAAGAGGACGACTGG 2702  
 Db 317 TrpAlaLysIleThrProLysHisSerAlaGlnAsnAsnSerPheArgAspGlnPhe 336  
 QY 2703 CAGAGCGCACAGCTGAGATCGTCCCATCGACGCCCTCCAC-----TTC 2747  
 Db 337 GlyArgLeuGlnThrGluThrValAlaIleAspAlaValArgAsnAlaGlyThrProLeu 356  
 QY 2748 AGACGCTACCTCGACCGAGTTTGTGCCGAGAGATCAGACGGGAGCTTAACAGAGCTTAC 2807  
 Db 357 GluCysLeuLeuAsnGlnLeuThrThrGluLysLeuThrArgGluValArgLysAlaAla 376  
 QY 2808 TGTGGATTCTTCCTCGTGGAGTTTCTTCAGAGAACCTGTCTGCAGTGGCTACAGGAAC 2867  
 Db 377 IleGlyPheLeuSerAlaGlyAspGlyPheSerLysIle---ProValValSerGlyTyr 395  
 QY 2868 TGGGGCTGTGTGCTTTGGGGGTGATGCTAGACTTAAGCCTTAATACAGATCCTGCA 2927  
 Db 396 TrpGlyCysGlyAlaPheArgLysAsnLysProLysLysPheLeuIleGlnValIleAla 415  
 QY 2928 GCTGCTGTAGTCGCGGACGACGTGTTTATTTTACCTTTGGGGACTCAGAACTGATGAGA 2987

Thu May 27 09:55:53 2004

```

Db      416 CysGlyIleSerAspArgProLeuGlnPheCysThrPheGlyAspThrGluLeuAlaLys 435
QY      2988 GACATTTCAGCATGCATACATTCCTCACTGAGAGAACTGACTGTTGGAGAGATATAT 3047
Db      436 LysCysGluGluMetThrLeuPheArgAsnAsnValArgThrGlyGlnLeuPhe 455
QY      3048 AAGCTGCTGCTAGCATATTACATGAAGATGCAGAACTGCTCCACCCCGGACCAGAC 3107
Db      456 LeuIleIle-----AsnSerIleGlyProPro 464
QY      3108 ATCAAGCTT---TATCCATTTCATATACCATGTCAGTTGCTCCTGTACACAGACCACC 3161
Db      465 LeuAsnTyrSerGluGlnTyrValPheAspAlaIleArgAlaLysIleAsnSerThr 483

```

Search completed: May 26, 2004, 17:07:11  
 Job time : 282.999 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 26, 2004, 18:38:05 ; Search time 31.3065 Seconds  
(without alignments)  
1611.120 Million cell updates/sec

Title: US-09-302-812-2  
Perfect score: 5184  
Sequence: 1 MSAGGCECTKRPWDAAA.....YHAVESCTQTNPQORTGA 977

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents AA:\*  
1: /cgn2\_6/prodata/2/iaa/5A COMB pep.\*  
2: /cgn2\_6/prodata/2/iaa/5B COMB pep.\*  
3: /cgn2\_6/prodata/2/iaa/6A COMB pep.\*  
4: /cgn2\_6/prodata/2/iaa/6B COMB pep.\*  
5: /cgn2\_6/prodata/2/iaa/PTUS COMB pep.\*  
6: /cgn2\_6/prodata/2/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	ID	Description
1	5184	100.0	977	4	US-09-302-812-2
2	5184	100.0	977	4	US-09-511-477-2
3	5184	100.0	977	4	US-09-511-507-2
4	4635.5	89.4	976	4	US-09-302-812-4
5	4635.5	89.4	976	4	US-09-511-477-4
6	4635.5	89.4	976	4	US-09-511-507-4
7	4332.5	83.6	968	4	US-09-302-812-6
8	4332.5	83.6	968	4	US-09-511-477-6
9	4332.5	83.6	968	4	US-09-511-507-6
10	1063.5	20.5	768	4	US-09-302-812-8
11	1063.5	20.5	768	4	US-09-511-477-8
12	1063.5	20.5	768	4	US-09-511-507-8
13	465	9.0	726	4	US-09-302-812-10
14	465	9.0	726	4	US-09-511-477-10
15	465	9.0	726	4	US-09-511-507-10
16	167	3.2	31	4	US-09-511-477-11
17	167	3.2	31	4	US-09-511-477-11
18	167	3.2	31	4	US-09-511-507-11
19	155.5	3.0	1804	4	US-09-362-336A-4
20	153	3.0	913	4	US-08-971-089-4
21	152	2.9	914	1	US-08-484-105-2
22	152	2.9	914	1	US-08-484-106-2
23	149	2.9	29	4	US-09-302-812-12
24	149	2.9	29	4	US-09-511-477-12
25	149	2.9	29	4	US-09-511-507-12
26	147	2.8	26	4	US-09-302-812-19
27	147	2.8	26	4	US-09-511-477-19

28	147	2.8	26	4	US-09-511-507-19	Sequence 19, Appl
29	146.5	2.8	1187	1	US-08-320-559-28	Sequence 28, Appl
30	146.5	2.8	1187	3	US-08-545-860D-28	Sequence 28, Appl
31	146.5	2.8	1187	5	PCT-US94-04496-28	Sequence 28, Appl
32	146.5	2.8	1210	1	US-08-320-559-26	Sequence 26, Appl
33	146.5	2.8	1210	3	US-08-545-860D-26	Sequence 26, Appl
34	146.5	2.8	1210	5	PCT-US94-04496-26	Sequence 26, Appl
35	143.5	2.8	717	3	US-08-910-925-1	Sequence 1, Appl
36	143.5	2.8	1177	4	US-09-134-001C-5106	Sequence 5106, Ap
37	143	2.8	703	3	US-08-910-925-4	Sequence 4, Appl
38	143	2.8	1783	4	US-09-362-336A-2	Sequence 2, Appl
39	142.5	2.7	1312	4	US-09-345-882-29	Sequence 29, Appl
40	141.5	2.7	1805	1	US-07-853-913-2	Sequence 2, Appl
41	141	2.7	778	6	5198347-4	Patent No. 5198347
42	140.5	2.7	951	4	US-09-125-635-8	Sequence 8, Appl
43	140.5	2.7	1420	4	US-09-125-635-4	Sequence 4, Appl
44	139.5	2.7	10182	4	US-09-134-001C-3159	Sequence 3159, Ap
45	137.5	2.7	879	4	US-09-914-259-38	Sequence 38, Appl

ALIGNMENTS

RESULT 1  
US-09-302-812-2  
; Sequence 2, Application US/09302812B  
; Patent No. 6333148  
; GENERAL INFORMATION:  
; APPLICANT: JACOBSON, Myron K.  
; APPLICANT: JACOBSON, Elaine L.  
; APPLICANT: AM, Jean-Christophe  
; APPLICANT: LIN, Winston  
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY(ADP-RIBOSE) GLYCOHYDROLASE (PARG) ENZY  
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIVE  
; TITLE OF INVENTION: THEREWITH  
; FILE REFERENCE: NIAD 201  
; CURRENT APPLICATION NUMBER: US/09/302,812B  
; CURRENT FILING DATE: 1999-04-30  
; EARLIER APPLICATION NUMBER: 60/083,768  
; EARLIER FILING DATE: 1998-05-01  
; NUMBER OF SEQ ID NOS: 38  
; SEQ ID NO 2  
; LENGTH: 977  
; TYPE: PRT  
; ORGANISM: Bos taurus  
; FEATURE:  
US-09-302-812-2

Query Match	100.0%;	Score 5184;	DB 4;	Length 977;
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches 977;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	MSAGGCECTKRPWDAAATSPPAASDARSPPGRRVLDSDKADPVQVRVPPSSSGCAL	60	
Db	1	MSAGGCECTKRPWDAAATSPPAASDARSPPGRRVLDSDKADPVQVRVPPSSSGCAL	60	
QY	61	GRAGQHRGATSIVFKQKITTSMDTKGKITVSESLHSESKENNTREESMMSSVQKDNFY	120	
Db	61	GRAGQHRGATSIVFKQKITTSMDTKGKITVSESLHSESKENNTREESMMSSVQKDNFY	120	
QY	121	QHNMEKLENVSQLGFDKSPVEKGTQYLKHQHTAMCKWQNEGPHSRLLESEPPATLVP	180	
Db	121	QHNMEKLENVSQLGFDKSPVEKGTQYLKHQHTAMCKWQNEGPHSRLLESEPPATLVP	180	
QY	181	EQFSNANVDQSSPKDDHSDTNSSESDNQOFLTHVKLANAKOTMEDEQGRARSHQCKG	240	
Db	181	EQFSNANVDQSSPKDDHSDTNSSESDNQOFLTHVKLANAKOTMEDEQGRARSHQCKG	240	
QY	241	ACHPAEACACQCEETDVVSESLDTSGETDVTGLKNANRLNROESSLGNPPFEKSE	300	
Db	241	ACHPAEACACQCEETDVVSESLDTSGETDVTGLKNANRLNROESSLGNPPFEKSE	300	
QY	301	PESPMVDVNSKNSCQDSEADETSPGDEQEDSSSAQTANKPSRFQFQREADTELKXSSA	360	

Db 301 PESPMVDVNSKNSQDSEADEETSPGFDEQEDSSSAQTANKPRFPQPREADTELKSSA 360  
QY 361 KGGEIRLHFQEGGESRAGNDVNAKPGSTSSLNVECRNSKQGRKDSKITDHFMRVPK 420  
Db 361 KGGEIRLHFQEGGESRAGNDVNAKPGSTSSLNVECRNSKQGRKDSKITDHFMRVPK 420  
QY 421 AEDKRKEQCEMKQRTKIPKIPPHLSDDKWLGTPIEEMRMPCGIRLPLPSAN 480  
Db 421 AEDKRKEQCEMKQRTKIPKIPPHLSDDKWLGTPIEEMRMPCGIRLPLPSAN 480  
QY 481 HTVTIRVDLLRIGEVKPPFTPHFKDLMDNKHVMPCEQNLYPVEDENGERRAAGSRWELI 540  
Db 481 HTVTIRVDLLRIGEVKPPFTPHFKDLMDNKHVMPCEQNLYPVEDENGERRAAGSRWELI 540  
QY 541 QTALLNRLTRPQNLKDALIKYNVAYSKKWDFLTALIDFWDKVLBEAQAHLIYQSILPDMVK 600  
Db 541 QTALLNRLTRPQNLKDALIKYNVAYSKKWDFLTALIDFWDKVLBEAQAHLIYQSILPDMVK 600  
QY 601 IALCLPNICTQPIPLLKQKNHSITMSQEQIASILLANAFCTPPRRNAKMKSEYSSYPDI 660  
Db 601 IALCLPNICTQPIPLLKQKNHSITMSQEQIASILLANAFCTPPRRNAKMKSEYSSYPDI 660  
QY 661 NFNLFEGRSSRKPEKLTLCYFPRRVTEKKPTGLVTFTRQSLEDFFEWERCEKLLTRH 720  
Db 661 NFNLFEGRSSRKPEKLTLCYFPRRVTEKKPTGLVTFTRQSLEDFFEWERCEKLLTRH 720  
QY 721 VTYEGTIEGNGQGMLOVDFAFNRVGGGVTSAGLVOEIRFLINPELIVSRLFTVLHDNE 780  
Db 721 VTYEGTIEGNGQGMLOVDFAFNRVGGGVTSAGLVOEIRFLINPELIVSRLFTVLHDNE 780  
QY 781 CLIIITGEQSEYTGAEYTRWASHEDRSERDDQWRTTEIVAIDALHFRYLDQFVPE 840  
Db 781 CLIIITGEQSEYTGAEYTRWASHEDRSERDDQWRTTEIVAIDALHFRYLDQFVPE 840  
QY 841 KIRRELKAYCGFLRPGVSSNLSAVATGNWCGAFGGDARLKALIQILAAVAERDVVY 900  
Db 841 KIRRELKAYCGFLRPGVSSNLSAVATGNWCGAFGGDARLKALIQILAAVAERDVVY 900  
QY 901 FTFGDSELMRDIYSMHTFLTERKLTVGVEYKLLRYNNEECRCNCSTPGPDIKLYPFIYHA 960  
Db 901 FTFGDSELMRDIYSMHTFLTERKLTVGVEYKLLRYNNEECRCNCSTPGPDIKLYPFIYHA 960  
QY 961 VESCTQTTNQPGQRTGA 977  
Db 961 VESCTQTTNQPGQRTGA 977

RESULT 2  
US-09-511-477-2  
; Sequence 2, Application US/09511477  
; Patent No. 6337202  
; GENERAL INFORMATION:  
; APPLICANT: JACOBSON, Myron K.  
; APPLICANT: JACOBSON, Elaine L.  
; APPLICANT: AME, Jean-Christophe  
; APPLICANT: LIN, Winston  
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE (PARG) EN  
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV  
; TITLE OF INVENTION: THEREWITH  
; FILE REFERENCE: NIAD 201  
; CURRENT APPLICATION NUMBER: US/09/511,477  
; CURRENT FILING DATE: 2000-02-23  
; PRIOR APPLICATION NUMBER: 09/302,812  
; PRIOR FILING DATE: 1999-04-30  
; NUMBER OF SEQ ID NOS: 38  
; SEQ ID NO 2  
; LENGTH: 977  
; TYPE: PRT  
; ORGANISM: Bos taurus  
; FEATURE:  
US-09-511-477-2

Query Match 100.0%; Score 5184; DB 4; Length 977;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 977; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MSAGPCCECTCKRPRWDAATSPPAASDARSPPGRORRVLDSDKADPVQVRVPVSSSSGAL 60  
Db 1 MSAGPCCECTCKRPRWDAATSPPAASDARSPPGRORRVLDSDKADPVQVRVPVSSSSGAL 60  
QY 61 GRAGQHRGATSILVFQKQITTSMDTKGIKTVSESLSHSENNNTREESMMSSVQKDNFY 120  
Db 61 GRAGQHRGATSILVFQKQITTSMDTKGIKTVSESLSHSENNNTREESMMSSVQKDNFY 120  
QY 121 QHNMEKLVNSQLGFDKSPVEKGTQYLKQHTAAMCKWQNEGPHSRLLSESEPPATLVP 180  
Db 121 QHNMEKLVNSQLGFDKSPVEKGTQYLKQHTAAMCKWQNEGPHSRLLSESEPPATLVP 180  
QY 181 EQFSNANVDQSSPKDDHSDTNSSESDNOQFLTHVLANAKQTMEDQEGREASHQCKG 240  
Db 181 EQFSNANVDQSSPKDDHSDTNSSESDNOQFLTHVLANAKQTMEDQEGREASHQCKG 240  
QY 241 ACHPAEACAGCQEEEDVVSSESPLSDTGSSEDTGTLKNNLRNROESSLGNSPFKESE 300  
Db 241 ACHPAEACAGCQEEEDVVSSESPLSDTGSSEDTGTLKNNLRNROESSLGNSPFKESE 300  
QY 301 PESPMVDVNSKNSQDSEADEETSPGFDEQEDSSSAQTANKPRFPQPREADTELKSSA 360  
Db 301 PESPMVDVNSKNSQDSEADEETSPGFDEQEDSSSAQTANKPRFPQPREADTELKSSA 360  
QY 361 KGGEIRLHFQEGGESRAGNDVNAKPGSTSSLNVECRNSKQGRKDSKITDHFMRVPK 420  
Db 361 KGGEIRLHFQEGGESRAGNDVNAKPGSTSSLNVECRNSKQGRKDSKITDHFMRVPK 420  
QY 421 AEDKRKEQCEMKQRTKIPKIPPHLSDDKWLGTPIEEMRMPCGIRLPLPSAN 480  
Db 421 AEDKRKEQCEMKQRTKIPKIPPHLSDDKWLGTPIEEMRMPCGIRLPLPSAN 480  
QY 481 HTVTIRVDLLRIGEVKPPFTPHFKDLMDNKHVMPCEQNLYPVEDENGERRAAGSRWELI 540  
Db 481 HTVTIRVDLLRIGEVKPPFTPHFKDLMDNKHVMPCEQNLYPVEDENGERRAAGSRWELI 540  
QY 541 QTALLNRLTRPQNLKDALIKYNVAYSKKWDFLTALIDFWDKVLBEAQAHLIYQSILPDMVK 600  
Db 541 QTALLNRLTRPQNLKDALIKYNVAYSKKWDFLTALIDFWDKVLBEAQAHLIYQSILPDMVK 600  
QY 601 IALCLPNICTQPIPLLKQKNHSITMSQEQIASILLANAFCTPPRRNAKMKSEYSSYPDI 660  
Db 601 IALCLPNICTQPIPLLKQKNHSITMSQEQIASILLANAFCTPPRRNAKMKSEYSSYPDI 660  
QY 661 NFNLFEGRSSRKPEKLTLCYFPRRVTEKKPTGLVTFTRQSLEDFFEWERCEKLLTRH 720  
Db 661 NFNLFEGRSSRKPEKLTLCYFPRRVTEKKPTGLVTFTRQSLEDFFEWERCEKLLTRH 720  
QY 721 VTYEGTIEGNGQGMLOVDFAFNRVGGGVTSAGLVOEIRFLINPELIVSRLFTVLHDNE 780  
Db 721 VTYEGTIEGNGQGMLOVDFAFNRVGGGVTSAGLVOEIRFLINPELIVSRLFTVLHDNE 780  
QY 781 CLIIITGEQSEYTGAEYTRWASHEDRSERDDQWRTTEIVAIDALHFRYLDQFVPE 840  
Db 781 CLIIITGEQSEYTGAEYTRWASHEDRSERDDQWRTTEIVAIDALHFRYLDQFVPE 840  
QY 841 KIRRELKAYCGFLRPGVSSNLSAVATGNWCGAFGGDARLKALIQILAAVAERDVVY 900  
Db 841 KIRRELKAYCGFLRPGVSSNLSAVATGNWCGAFGGDARLKALIQILAAVAERDVVY 900  
QY 901 FTFGDSELMRDIYSMHTFLTERKLTVGVEYKLLRYNNEECRCNCSTPGPDIKLYPFIYHA 960  
Db 901 FTFGDSELMRDIYSMHTFLTERKLTVGVEYKLLRYNNEECRCNCSTPGPDIKLYPFIYHA 960  
QY 961 VESCTQTTNQPGQRTGA 977  
Db 961 VESCTQTTNQPGQRTGA 977

RESULT 3  
 US-09-511-507-2  
 ; Sequence 2, Application US/09511507  
 ; Patent No. 6395543  
 ; GENERAL INFORMATION:  
 ; APPLICANT: JACOBSON, Myron K.  
 ; APPLICANT: JACOBSON, Elaine L.  
 ; APPLICANT: AME, Jean-Christophe  
 ; APPLICANT: LIN, Winston  
 ; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE (PARG) EN  
 ; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV  
 ; TITLE OF INVENTION: THEREWITH  
 ; CURRENT APPLICATION NUMBER: US/09/511,507  
 ; CURRENT FILING DATE: 2000-02-23  
 ; PRIOR APPLICATION NUMBER: 09/302,812  
 ; PRIOR FILING DATE: 1999-04-30  
 ; NUMBER OF SEQ ID NOS: 38  
 ; SEQ ID NO 2  
 ; LENGTH: 977  
 ; TYPE: PRT  
 ; ORGANISM: Bos taurus  
 ; FEATURE:  
 US-09-511-507-2  
 Query Match 100.0%; Score 5184; DB 4; Length 977;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 977; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MSAGPCECTKPRWDAAATSPPAASDARSPPGRRVLDSDKADPQVRVPPSSGGCAL 60  
 DB 1 MSAGPCECTKPRWDAAATSPPAASDARSPPGRRVLDSDKADPQVRVPPSSGGCAL 60  
 QY 61 GRAGQHRGATSILVFKQKITSMDTKGKTVESESLHSEKNNNTREESMSSVQKDNFY 120  
 DB 61 GRAGQHRGATSILVFKQKITSMDTKGKTVESESLHSEKNNNTREESMSSVQKDNFY 120  
 QY 121 QHNMEKLVNSQLGFDKSPVEKGTQYLKQHTAAMCKQWNEGPHSRLLSEPPAVTLVP 180  
 DB 121 QHNMEKLVNSQLGFDKSPVEKGTQYLKQHTAAMCKQWNEGPHSRLLSEPPAVTLVP 180  
 QY 181 EQFSNANVDQSSPKDHDSDTNSSESDNQFLTHVKLANAKQTMEDQEGREARSHOKCKG 240  
 DB 181 EQFSNANVDQSSPKDHDSDTNSSESDNQFLTHVKLANAKQTMEDQEGREARSHOKCKG 240  
 QY 241 ACHPAACAGCQCEETDVVSESLSDTGSDEDTGLKNANLNROBSSSLGNSPPFEKSE 300  
 DB 241 ACHPAACAGCQCEETDVVSESLSDTGSDEDTGLKNANLNROBSSSLGNSPPFEKSE 300  
 QY 301 PESPMVDVNSKNSCQDSEADEETSPGDEQEDSSAQNTANKPQRPQPREADTELKRS 360  
 DB 301 PESPMVDVNSKNSCQDSEADEETSPGDEQEDSSAQNTANKPQRPQPREADTELKRS 360  
 QY 361 KGGEIPLHFQEGESRAGNDVNAKPGSTSLNVECRNSKOHGRKDSKITDHFMRVVK 420  
 DB 361 KGGEIPLHFQEGESRAGNDVNAKPGSTSLNVECRNSKOHGRKDSKITDHFMRVVK 420  
 QY 421 AEDKRKEQCEMKHQRTERKIPKIPPHLSDPDKWLGTPPEEMRMRPCGIRLPLRPSAN 480  
 DB 421 AEDKRKEQCEMKHQRTERKIPKIPPHLSDPDKWLGTPPEEMRMRPCGIRLPLRPSAN 480  
 QY 481 HTVTIRVDLLRIGEVKPPFTPHFKDLMDNKHVKMPCSEQNLVPEVDENGERRAGRWELI 540  
 DB 481 HTVTIRVDLLRIGEVKPPFTPHFKDLMDNKHVKMPCSEQNLVPEVDENGERRAGRWELI 540  
 QY 541 QTALLNRLTRPNLKDAILKYNVAYSKKMDFTALIDFMDKVLKEAAQHLVQSILPDMVK 600  
 DB 541 QTALLNRLTRPNLKDAILKYNVAYSKKMDFTALIDFMDKVLKEAAQHLVQSILPDMVK 600  
 QY 601 IALCLPNICTQPIPLLKQNMHSITMSQEQIASILLANAFCTPPRRNAKMKSEYSSYPDI 660  
 DB 601 IALCLPNICTQPIPLLKQNMHSITMSQEQIASILLANAFCTPPRRNAKMKSEYSSYPDI 660

QY 661 NFNLFEGRSRKPEKLTLCYFERRVTEKPTGLVTFTRQSLDFFPEWCEKLLTRLH 720  
 DB 661 NFNLFEGRSRKPEKLTLCYFERRVTEKPTGLVTFTRQSLDFFPEWCEKLLTRLH 720  
 QY 721 VTYEGTIEGNGQGMQLQVDFANRFVGGVTSAGLVQEEIRFLINPELIVSRFLTVEVDHNE 780  
 DB 721 VTYEGTIEGNGQGMQLQVDFANRFVGGVTSAGLVQEEIRFLINPELIVSRFLTVEVDHNE 780  
 QY 781 CLITGTIOYSEYTGAYTYWASHEDSRDDWQRTTEIVALDHLHFRYLDQFVPE 840  
 DB 781 CLITGTIOYSEYTGAYTYWASHEDSRDDWQRTTEIVALDHLHFRYLDQFVPE 840  
 QY 841 KIRRELKAYCGFLRPGVSSNLSAVATGNWCGAFGGDARLKALIQILAAVAERDVVY 900  
 DB 841 KIRRELKAYCGFLRPGVSSNLSAVATGNWCGAFGGDARLKALIQILAAVAERDVVY 900  
 QY 901 FTFGDSSELMRDIYSMHTFLTERTKLTGVRVYKLLAYNEECNCSTPGPDIKLYPFIYHA 960  
 DB 901 FTFGDSSELMRDIYSMHTFLTERTKLTGVRVYKLLAYNEECNCSTPGPDIKLYPFIYHA 960  
 QY 961 VESCTQTNNQPGQTGA 977  
 DB 961 VESCTQTNNQPGQTGA 977  
 RESULT 4  
 US-09-302-812-4  
 ; Sequence 4, Application US/09302812B  
 ; Patent No. 6333148  
 ; GENERAL INFORMATION:  
 ; APPLICANT: JACOBSON, Myron K.  
 ; APPLICANT: JACOBSON, Elaine L.  
 ; APPLICANT: AM, Jean-Christophe  
 ; APPLICANT: LIN, Winston  
 ; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE (PARG) ENZ  
 ; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIVE  
 ; TITLE OF INVENTION: THEREWITH  
 ; FILE REFERENCE: NIAD 201  
 ; CURRENT APPLICATION NUMBER: US/09/302,812B  
 ; CURRENT FILING DATE: 1999-04-30  
 ; EARLIER APPLICATION NUMBER: 60/083,768  
 ; EARLIER FILING DATE: 1998-05-01  
 ; NUMBER OF SEQ ID NOS: 38  
 ; SEQ ID NO 4  
 ; LENGTH: 976  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 ; FEATURE:  
 US-09-302-812-4  
 Query Match 89.4%; Score 4635.5; DB 4; Length 976;  
 Best Local Similarity 89.5%; Pred. No. 0;  
 Matches 874; Conservative 32; Mismatches 69; Indels 1; Gaps 1;  
 QY 1 MSAGPCECTKPRWDAAATSPPAASDARSPPGRRVLDSDKADPQVRVPPSSGGCAL 60  
 DB 1 MSAGPCECTKPRWDAAATSPPAASDARSPPGRRVLDSDKADPQVRVPPSSGGCAL 60  
 QY 61 GRAGQHRGATSILVFKQKITSMDTKGKTVESESLHSEKNNNTREESMSSVQKDNFY 120  
 DB 61 GRAGQHRGATSILVFKQKITSMDTKGKTVESESLHSEKNNNTREESMSSVQKDNFY 120  
 QY 121 QHNMEKLVNSQLGFDKSPVEKGTQYLKQHTAAMCKQWNEGPHSRLLSEPPAVTLVP 180  
 DB 121 QHNMEKLVNSQLGFDKSPVEKGTQYLKQHTAAMCKQWNEGPHSRLLSEPPAVTLVP 180  
 QY 181 EQFSNANVDQSSPKDHDSDTNSSESDNQFLTHVKLANAKQTMEDQEGREARSHOKCKG 240  
 DB 181 EQFSNANVDQSSPKDHDSDTNSSESDNQFLTHVKLANAKQTMEDQEGREARSHOKCKG 240  
 QY 241 ACHPAACAGCQCEETDVVSESLSDTGSDEDTGLKNANLNROBSSSLGNSPPFEKSE 300  
 DB 241 ACHPAACAGCQCEETDVVSESLSDTGSDEDTGLKNANLNROBSSSLGNSPPFEKSE 300  
 QY 301 PESPMVDVNSKNSCQDSEADEETSPGDEQEDSSAQNTANKPQRPQPREADTELKRS 360  
 DB 301 PESPMVDVNSKNSCQDSEADEETSPGDEQEDSSAQNTANKPQRPQPREADTELKRS 360  
 QY 361 KGGEIPLHFQEGESRAGNDVNAKPGSTSLNVECRNSKOHGRKDSKITDHFMRVVK 420  
 DB 361 KGGEIPLHFQEGESRAGNDVNAKPGSTSLNVECRNSKOHGRKDSKITDHFMRVVK 420  
 QY 421 AEDKRKEQCEMKHQRTERKIPKIPPHLSDPDKWLGTPPEEMRMRPCGIRLPLRPSAN 480  
 DB 421 AEDKRKEQCEMKHQRTERKIPKIPPHLSDPDKWLGTPPEEMRMRPCGIRLPLRPSAN 480  
 QY 481 HTVTIRVDLLRIGEVKPPFTPHFKDLMDNKHVKMPCSEQNLVPEVDENGERRAGRWELI 540  
 DB 481 HTVTIRVDLLRIGEVKPPFTPHFKDLMDNKHVKMPCSEQNLVPEVDENGERRAGRWELI 540  
 QY 541 QTALLNRLTRPNLKDAILKYNVAYSKKMDFTALIDFMDKVLKEAAQHLVQSILPDMVK 600  
 DB 541 QTALLNRLTRPNLKDAILKYNVAYSKKMDFTALIDFMDKVLKEAAQHLVQSILPDMVK 600  
 QY 601 IALCLPNICTQPIPLLKQNMHSITMSQEQIASILLANAFCTPPRRNAKMKSEYSSYPDI 660  
 DB 601 IALCLPNICTQPIPLLKQNMHSITMSQEQIASILLANAFCTPPRRNAKMKSEYSSYPDI 660

```

301 PESMDVNDKNSQDSEADSETSPGDEQEDSSAQTANKPSRFPQREADTELKRSXA 360
300 PESMDVNDKNSQDSEADSETSPGDEQEDSSAQTANKPSRFPQREADTELKRSXA 359
361 KGGIRLHFQEGGESRAGMNDVNAKPGSTSLNVECRNSKQGRKDSKITDHFVRVPK 420
360 KGGIRLHFQEGGESRAGMNDVNAKPGSTSLNVECRNSKQGRKDSKITDHFVRVPK 419
421 AEDKKEQCEMCHORTERKIPKYPHLSPPDKWLTPIEMRMRPCGIRLPLRPSAN 480
420 AEDKKEQCEMCHORTERKIPKYPHLSPPDKWLTPIEMRMRPCGIRLPLRPSAN 479
481 HTVIRVDLLRIGVVPKPFTHFKDLMDNKHVMKPCSEONLYPVEDENGERRAAGSRWELI 540
480 HTVIRVDLLRIGVVPKPFTHFKDLMDNKHVMKPCSEONLYPVEDENGERRAAGSRWELI 539
541 QTALLNRLTRPQNLKDALIKYNVAYSKKWDFTALIDFMDKVLBEAQAHLVOSILPDMVK 600
540 QTALLNRLTRPQNLKDALIKYNVAYSKKWDFTALIDFMDKVLBEAQAHLVOSILPDMVK 599
601 IALCLPNICTQPIPLKQKMHNSITMSQEQIASLLANAFCTFPRRNAMKSEYSSYPDI 660
600 IALCLPNICTQPIPLKQKMHNSITMSQEQIASLLANAFCTFPRRNAMKSEYSSYPDI 659
661 NFNLFEGRSSRKPEKLTLCYFRRVTEKKPTGLVTFTRQSLDDFPWERCERKLTSLH 720
660 NFNLFEGRSSRKPEKLTLCYFRRVTEKKPTGLVTFTRQSLDDFPWERCERKLTSLH 719
721 VTYEGTIEGNGQMLQVDFANRFGGGVTSAGLVOEEIRFLINPELIVSRLEFTEVDHNE 780
720 VTYEGTIEGNGQMLQVDFANRFGGGVTSAGLVOEEIRFLINPELIVSRLEFTEVDHNE 779
781 CLITGTQYSEYTGAYETYSRWSHEDRSERDDQWRTTEIVAIDALHFRYLDQFVPE 840
780 CLITGTQYSEYTGAYETYSRWSHEDRSERDDQWRTTEIVAIDALHFRYLDQFVPE 839
841 KIRRELKAYCGFLRPGVSSNLSAVATGNMGCGAFGGDARLKALIQILAAVAERDVVY 900
840 KIRRELKAYCGFLRPGVSSNLSAVATGNMGCGAFGGDARLKALIQILAAVAERDVVY 899
901 FTFGDSELMRDIYSMHTLTERKLTGVGVYKLLRYNNEECNCSCTPGDILKLYFIYHA 960
900 FTFGDSELMRDIYSMHTLTERKLTGVGVYKLLRYNNEECNCSCTPGDILKLYFIYHA 959

961 VESCTQTTNQPQRTG 976
960 VESCAETADHSQRTG 975

RESULT 5
US-09-511-477-4
; Sequence 4, Application US/09511477
; Patent No. 6337202
; GENERAL INFORMATION:
; APPLICANT: JACOBSON, Myron K.
; APPLICANT: JACOBSON, Elaine L.
; APPLICANT: AME, Jean-Christophe
; APPLICANT: LIN, Winston
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOPOLYMERASE (PARG) EN
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV
; FILE REFERENCE: NIAD 201
; CURRENT FILING DATE: 2000-02-23
; PRIOR APPLICATION NUMBER: 09/302,812
; PRIOR FILING DATE: 1999-04-30
; NUMBER OF SEQ ID NOS: 38
; SEQ ID NO 4
; LENGTH: 976
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:

```

```

US-09-511-477-4
Query Match      89.4%; Score 4635.5; DB 4; Length 976;
Best Local Similarity 89.5%; Pred. No. 0;
Matches 874; Conservative 32; Mismatches 69; Indels 1; Gaps 1;

QY 1 MSAGPCGCECTKPRWDAATSPPAASDARSFPGRRVLDSDKADPVQFRVPPSSSSGAL 60
DB 1 MNAGPCGCECTKATRGAAATTS-PAASDARSFPGRRVLDSDKADPVQFRVPPSSSPACVP 59
61 GRAGQHRGSATSILVFKQKITTSWMDTKGKTKVTSSESLHSENNNTREESMMSSVQKDNFY 120
60 GQAGQHRGSATSILVFKQKITTSWMDTKGKTKVTSSESLHSENNNTREESMMSSVQKDNFY 119
121 QHNEKLENTVSLQFDPKSPVEKGTQYILKQHOTAMCKWQNEGSEHSELESEPPAVTLVP 180
120 QHNEKLENTVSLQFDPKSPVEKGTQYILKQHOTAMCKWQNEGSEHSELESEPPAVTLVP 179
181 EQFSNANVDQSSPKDDHSTNSESBDNQFLTHVKLANAKQTMEDQGRKSHQCKGK 240
180 EQFSNANVDQSSPKDDHSTNSESBDNQFLTHVKLANAKQTMEDQGRKSHQCKGK 239
241 ACHPAEACQCEQETDVVSESLDSTGSDVGTGLKXANRLNROESSLGNSSPFKESE 300
240 SCHPGEDCASQCEQETDVVSESLDSTGSDVGTGLKXANRLNROESSLGNSSPFKESE 299
301 PESMDVNDKNSQDSEADSETSPGDEQEDSSAQTANKPSRFPQREADTELKRSXA 360
300 PESMDVNDKNSQDSEADSETSPGDEQEDSSAQTANKPSRFPQREADTELKRSXA 359
361 KGGIRLHFQEGGESRAGMNDVNAKPGSTSLNVECRNSKQGRKDSKITDHFVRVPK 420
360 KGGIRLHFQEGGESRAGMNDVNAKPGSTSLNVECRNSKQGRKDSKITDHFVRVPK 419
421 AEDKKEQCEMCHORTERKIPKYPHLSPPDKWLTPIEMRMRPCGIRLPLRPSAN 480
420 AEDKKEQCEMCHORTERKIPKYPHLSPPDKWLTPIEMRMRPCGIRLPLRPSAN 479
481 HTVIRVDLLRIGVVPKPFTHFKDLMDNKHVMKPCSEONLYPVEDENGERRAAGSRWELI 540
480 HTVIRVDLLRIGVVPKPFTHFKDLMDNKHVMKPCSEONLYPVEDENGERRAAGSRWELI 539
541 QTALLNRLTRPQNLKDALIKYNVAYSKKWDFTALIDFMDKVLBEAQAHLVOSILPDMVK 600
540 QTALLNRLTRPQNLKDALIKYNVAYSKKWDFTALIDFMDKVLBEAQAHLVOSILPDMVK 599
601 IALCLPNICTQPIPLKQKMHNSITMSQEQIASLLANAFCTFPRRNAMKSEYSSYPDI 660
600 IALCLPNICTQPIPLKQKMHNSITMSQEQIASLLANAFCTFPRRNAMKSEYSSYPDI 659
661 NFNLFEGRSSRKPEKLTLCYFRRVTEKKPTGLVTFTRQSLDDFPWERCERKLTSLH 720
660 NFNLFEGRSSRKPEKLTLCYFRRVTEKKPTGLVTFTRQSLDDFPWERCERKLTSLH 719
721 VTYEGTIEGNGQMLQVDFANRFGGGVTSAGLVOEEIRFLINPELIVSRLEFTEVDHNE 780
720 VTYEGTIEGNGQMLQVDFANRFGGGVTSAGLVOEEIRFLINPELIVSRLEFTEVDHNE 779
781 CLITGTQYSEYTGAYETYSRWSHEDRSERDDQWRTTEIVAIDALHFRYLDQFVPE 840
780 CLITGTQYSEYTGAYETYSRWSHEDRSERDDQWRTTEIVAIDALHFRYLDQFVPE 839
841 KIRRELKAYCGFLRPGVSSNLSAVATGNMGCGAFGGDARLKALIQILAAVAERDVVY 900
840 KIRRELKAYCGFLRPGVSSNLSAVATGNMGCGAFGGDARLKALIQILAAVAERDVVY 899
901 FTFGDSELMRDIYSMHTLTERKLTGVGVYKLLRYNNEECNCSCTPGDILKLYFIYHA 960
900 FTFGDSELMRDIYSMHTLTERKLTGVGVYKLLRYNNEECNCSCTPGDILKLYFIYHA 959

961 VESCTQTTNQPQRTG 976
960 VESCAETADHSQRTG 975

```

```

600 IALCLPNICTQPIILLKQKXNHSITMSQEQIASLLANAFFCTFPRRNAKMKSEYSSYPDI 659
661 NFNFLFEGRSSRKPEKLTFLCYFRRVTEKKPTGLVTFTROSLEDFFPWEKCEKLLTLHL 720
660 NFNFLFEGRSSRKPEKLTFLCYFRRVTEKKPTGLVTFTROSLEDFFPWEKCEKPLTLHL 719
721 VTYEGTIEGCGQMLQVDFANRFVGGGVTSAGLVQEEIRFLINPELIIVSRFLTVEVLHDNE 780
720 VTYEGTIEGCGQMLQVDFANRFVGGGVTSAGLVQEEIRFLINPELIIVSRFLTVEVLHDNE 779
781 CLIIITGTEQYSEYTGAYETVWARSHEDSRDDQWRETTEIVAIDALHFRYRLDQFVPE 840
780 CLIIITGTEQYSEYTGAYETVWARSHEDSRDDQWRETTEIVAIDALHFRYRLDQFVPE 839
841 KIRRELKAYCGFLRPGVSSNLSAVATGNWCGGAFGGDARLKALIQILAAVAERDVVY 900
840 KMRRELKAYCGFLRPGVSSNLSAVATGNWCGGAFGGDARLKALIQILAAVAERDVVY 899
901 FTFGDSELMDIYSWHITELTERKLTVGVEYKLLRLRYNEECNCSTPGDIKLYPFIYHA 960
900 FTFGDSELMDIYSWHITELTERKLTVGVEYKLLRLRYNEECNCSTPGDIKLYPFIYHA 959
961 VESCTQTITNQCQRTG 976
960 VESCAETADHSQRTG 975

RESULT 7
US-09-302-812-6
; Sequence 6, Application US/09302812B
; Patent No. 633148
; GENERAL INFORMATION:
; APPLICANT: JACOBSON, Myron K.
; APPLICANT: JACOBSON, Elaine L.
; APPLICANT: AM, Jean-Christophe
; APPLICANT: LIN, Winston
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE (PARG) ENZ
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIVE
; TITLE OF INVENTION: THEREWITH
; FILE REFERENCE: NIAD 201
; CURRENT APPLICATION NUMBER: US/09/302,812B
; CURRENT FILING DATE: 1999-04-30
; EARLIER APPLICATION NUMBER: 60/083,768
; EARLIER FILING DATE: 1998-05-01
; NUMBER OF SEQ ID NOS: 38
; SEQ ID NO 6
; LENGTH: 968
; TYPE: PRT
; ORGANISM: Mus musculus
; FEATURE:
US-09-302-812-6

Query Match 83.6%; Score 4332.5; DB 4; Length 968;
Best Local Similarity 84.0%; Pred. No. 0;
Matches 820; Conservative 57; Mismatches 90; Indels 9; Gaps 5;

QY 1 MSAGPGCEPCTKPRWDAATSPPAASDARSFPGRQRRVLDSDAPVQFRVPVSSSGCAL 60
DB 1 MSAGPGWEPTCK-ARWGAAGTSAPTASDSRSFPGRQRRVLDPKDAPVQFRVPVSSPACVS 59
QY 61 GRAGQHRGSATSIVFKQKTTSMQTKITVSESLHSKENNNTREESMMSSVQKDNFY 120
DB 60 GRAGPHRGNATSFVFKQKTTTMDTKGPKTAESE---SKENNNTRIDSMSSVQKDNFY 116
QY 121 QHNWEKLENTVSQIGFDKSPVEKCTOYLKQHOTAKMCKWQNEGPHSRLLSESEPPAVTLVP 180
DB 117 PHKVEKLENTVPLNLDKSPTEKSQYLNQOOTASVCKWQNEGPHSRLLSESEPPAGTFLP 176
QY 181 EQSNANVDQSSPKDHDSDTNSBESDNOQFLTHVKLANAKQTWEDEQGEARSHQCGK 240
DB 177 KQLSNANIGQSPHTDHDSDTDEEDRDNQOFLPIKLANTKPTVGDQ---ASSNCKCSG 233
QY 241 ACHPAEACAGCQGEETDVWSESFLSDTSGSEDTGTGLKNANRLNRQSSLGNSPPEKESE 300

```





Db 952 VESSAETDMPGQKAG 967  
||| :||: |||: |  
RESULT 9  
US-09-511-507-6  
; Sequence 6, Application US/09511507  
; Patent No. 6395543  
; GENERAL INFORMATION:  
; APPLICANT: JACOBSON, Myron K.  
; APPLICANT: JACOBSON, Elaine L.  
; APPLICANT: AWE, Jean-Christophe  
; APPLICANT: LIN, Winston  
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOSYLTRANSFERASE (PARG) EN  
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV  
; FILE REFERENCE: NIAD 201  
; CURRENT APPLICATION NUMBER: US/09/511,507  
; PRIOR FILING DATE: 2000-02-23  
; PRIOR FILING DATE: 1999-04-30  
; NUMBER OF SEQ ID NOS: 38  
; SEQ ID NO 6  
; LENGTH: 968  
; TYPE: PRT  
; ORGANISM: Mus musculus  
; FEATURE:  
US-09-511-507-6  
Query Match 83.6%; Score 4332.5; DB 4; Length 968;  
Best Local Similarity 84.0%; Pred. No. 0;  
Matches 820; Conservative 57; Mismatches 90; Indels 9; Gaps 5;  
QY 1 MSAGPGCEPCTKPRMDAAATSPPAASDARSFPGQRRLVDSKADAPVQFVPPVSSSGCAL 60  
Db 1 MSAGPGWEPCTK-ARWGAAGTSAPTASDSRSFPGQRRLVDPKADAPVQFVPPVSSSPACVS 59  
QY 61 GRAGQHGSAATSLVFKQTTTSMWDTKGIKTVSESLHSKNNTRSESNMSSVQKDNFY 120  
Db 60 GRAGPRGNATSVFVKQTTTSMWDTKGPKTASE---SKENNNTRIDSMSSVQKDNFY 116  
QY 121 QHNEKLENYSGQFDSKSPVEKGTQYLKQHTQAMCKWQNEGPHSEKLLSEPPAVTLVP 180  
Db 117 PHKVEKLENYQLNDLSPEKSSQYLNQQTASVCWQNEGKHAQQLLASEPPAGTPLP 176  
QY 181 EQFSNANVDQSSPKDHSNDSESRNQQLTHVKLANAKQTMEDQCGEARSHQKCGK 240  
Db 177 KQLSNANIGQSPHTDHSDDHEEDRDNDQFLTPIKLANTKPTVGDGQ---ARSNCKCSG 233  
QY 241 ACPAEACAGCQOQETOVVSEPLSDTGSDEVTGGLKNNRLNQESSLGNSSPPFEKSE 300  
Db 234 SRQSVKDCQCGQEEVDVLPESPLSDVGAEDIGTPKNDKLTQCESLSDGSPFEKSE 293  
QY 301 PESPMVDNSKNSQDSEADETSFGFDEQDSSSAQTANKPSRFPQPREADTELKRS 360  
Db 294 PESPMVDNSKNSQDSEADETSFVDEQDRESS-QTANKLSSCQAREADGLKRYLT 352  
QY 361 KGBIRLHFQEGESRAGMVDNNAKRPSTSSINVECRNSKQGRKDSKITDHPMRVPK 420  
Db 353 KGSVRLHFQPE-GENNAAGTSDLNKAPSGNSSLNVECRNSKQGRKDSKITDHPMRISK 411  
QY 421 AEDKKEQCEMKHQRTERKIPKYPHLSPPKMLGTPIEMRMPRCGIRLPLRPSAN 480  
Db 412 SEDRKEQCEVHQRTERKIPKYPNLPPEKMLGTPIEMRMPRCGIRLPLRPSAS 471  
QY 481 HTVTIRVDLLRIGEVPKFPFTHFKDLDNKNVXQMPCESEONLYPVEDENGERAAGSRWELI 540  
Db 472 HTVTIRVDLLRAGEVPKFPFTHFKDLDNKNVXQMPCESEONLYPVEDENGERAAGSRWELI 531  
QY 541 QTALNRLTRPNLKDAILKYNVASKWDFTALIDFMDKVLREAEAOHLVQSILPDMVK 600  
Db 532 QTALNKFTRPNLKDAILKYNVASKWDFTALIDFMDKVLREAEAOHLVQSILPDMVK 591

QY 601 TALCLPNICTQPIPLLKQKMHSHITMSQEQIASILLANAFCTFFPRRNAMKSESYSPDI 660  
Db 592 TALCLPNICTQPIPLLKQKMHSHITMSQEQIASILLANAFCTFFPRRNAMKSESYSPDI 651  
QY 661 NFNRLFEGRSSRKPCKLTKLFCYFRVTEKKPTGLVTFTRQSLDEFPWECEKLLTRLH 720  
Db 652 NFNRLFEGRSSRKPCKLTKLFCYFRVTEKKPTGLVTFTRQSLDEFPWECEKLLTRLH 711  
QY 721 VTYEGTIEGNGQMLQVDPANFVGGVTSAGLVQEEIRFLINPELIVSRLFTVLDHNE 780  
Db 712 VTYEGTIEGNGQMLQVDPANFVGGVTSAGLVQEEIRFLINPELIVSRLFTVLDHNE 771  
QY 781 CLIIITGEQSYTGYAETRWARSHEDSRDDWQRRTEIIVADLHFRYLDQFVE 840  
Db 772 CLIIITGEQSYTGYAETRWARSHEDSRDDWQRRTEIIVADLHFRYLDQFVE 831  
QY 841 KIRRELKAYCGFLRPGVSENLAVATGNWCGAFGGDARLKALQIILAAAIAERDVVY 900  
Db 832 KVRRELKAYCGFLRPGVSENLAVATGNWCGAFGGDARLKALQIILAAAIAERDVVY 891  
QY 901 FTFGDSELMRDIVSMHTFLTERKLTVEYVKLILRYNEECRNCSTPGPDIKLYPIYHA 960  
Db 892 FTFGDSELMRDIVSMHTFLTERKLTVEYVKLILRYNEECRNCSTPGPDIKLYPIYHA 951  
QY 961 VESCTOTTNQPGQRTG 976  
Db 952 VESSAETDMPGQKAG 967  
RESULT 10  
US-09-302-812-8  
; Sequence 8, Application US/09302812B  
; Patent No. 6333148  
; GENERAL INFORMATION:  
; APPLICANT: JACOBSON, Myron K.  
; APPLICANT: JACOBSON, Elaine L.  
; APPLICANT: AM, Jean-Christophe  
; APPLICANT: LIN, Winston  
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOSYLTRANSFERASE (PARG) ENZ  
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIVE  
; FILE REFERENCE: NIAD 201  
; CURRENT APPLICATION NUMBER: US/09/302,812B  
; CURRENT FILING DATE: 1999-04-30  
; EARLIER APPLICATION NUMBER: 60/083,768  
; EARLIER FILING DATE: 1998-05-01  
; NUMBER OF SEQ ID NOS: 38  
; SEQ ID NO 8  
; LENGTH: 768  
; TYPE: PRT  
; ORGANISM: Drosophila melanogaster  
; FEATURE:  
US-09-302-812-8  
Query Match 20.5%; Score 1063.5; DB 4; Length 768;  
Best Local Similarity 41.5%; Pred. No. 5.9e-90;  
Matches 244; Conservative 93; Mismatches 198; Indels 53; Gaps 15;  
QY 417 RVPKAEKKEQCEMKHQRTERKIPKYPHLS--SPDKMLGTPIERM--RRMRCGIRL 472  
Db 45 RMGKSPDGGISLETETEE-----PENLANSLLDSWRGVSMEALHNRNQPFLENL 94  
QY 473 PPLRPSANHTVTIRVDLLRIGEVPKFPFTHFKDLDNKNVXQMPCESEONLYPVEDENGERA 532  
Db 95 PPTVAGNLHRVMYQ---LPFIRETP-PRPKSPGKWDSEHVRVLPKAPESKYPRENPDGSTT 150  
QY 533 AGSRWELIQALNRLTRPNLKDAILKYNVASKWDFTALIDFMDKVLREAEAOHLVQ 592  
Db 151 IDPFWEMIERALLQPIKTCBELQAAIISYNTTYRDQWHFRALHQLDEEDSETRVFFE 210  
QY 593 SILPDMVKIALCLPNICTQPIPLLKQKMHSHITMSQEQIASILLANAFCTFFPRRNAMK 651  
Db 211 DLLPRTIRLRLPDLIQSPVPLKHKHNSLSUSQOISCLLANAFCTFFPRNTLKRK 270

Qy	652	SEYSSYDDINRNLPEGRSSRKPXKTLFCYFRV--TEKK-----PGLVTFTRQS--L	703
Db	271	SEYTFDDINRNLQYSGTGPVLEKKIMHYFRVRCPTERDASNVPTGVVTFVRSGLP	330
Qy	704	EDFPERCEKLL--TRLHVTYEGTIEGNGQMLQVDFANFRVGGVTSAGLVQBEIRPL	761
Db	331	EHLIDWSAAPLGDVPLHVDABGTIEDBGILLQVDFANKVLGGVLGHGCVQBEIRFV	390
Qy	762	INPLIVSRLTEVLDDHNECLITCTEOYSEVTVGYAETVWARSHEDRSERDDMORRTTE	821
Db	391	ICPELLVGLKTECURLPPEALVMLGAERISNTTGHAGFWSGCFEDSTPRDSSGRQTA	450
Qy	822	IVADALFRRLYLDQFVPEKIRRELNKAYCGFLR-----PGVSSENISAVATGNWGGA	875
Db	451	IVADALHFAQSHHQYQREDLMBRELNKAYIGEVHVMVTPPG-----VATGNWGGA	502
Qy	876	FGGDARLAKILQLAAVAARVDVYFTFGDSLMDIYSMHTFLTERKLTGVCEVKKLILR	935
Db	503	FGGUSYLKALIQLMVCAQLGRPLAYTTCNVFREDDFHEMWLLFRNDGTIVQOLWS--ILR	561
Qy	936	YVNEBCRNCSTGPP-----DIKLVPIFVHAVESCCTQTTNQBGQRGA	977
Db	562	YSRLRIEKSKEPRENKASKKKLYDFI---KEELKKVDRVPEGASA	606

RESULT 11  
US-09-511-477-8  
Sequence 8, Application US/09511477  
Patent No. 6337202  
GENERAL INFORMATION:  
APPLICANT: JACOBSON, Myron K.  
APPLICANT: JACOBSON, Elaine L.  
APPLICANT: AME, Jean-Christophe  
APPLICANT: LIN, Winston  
TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE (PARG) ENZY  
TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV  
TITLE OF INVENTION: THEREWITH  
FILE REFERENCE: NIAID 201  
CURRENT APPLICATION NUMBER: US/09/511,477  
CURRENT FILING DATE: 2000-02-23  
PRIOR APPLICATION NUMBER: 09/302,812  
PRIOR FILING DATE: 1999-04-30  
NUMBER OF SEQ ID NOS: 38  
SEQ ID NO 8  
LENGTH: 768  
TYPE: PRT  
ORGANISM: Drosophila melanogaster  
FEATURE:  
US-09-511-477-8

[illegible]

```

RESULT 12
US-09-511-507-8
; Sequence 8, Application US/09511507
; Patent No. 6395543
; GENERAL INFORMATION:
; APPLICANT: JACOBSON, Myron K.
; APPLICANT: JACOBSON, Elaine J.
; APPLICANT: AME, Jean-Christophe
; APPLICANT: LIN, Winston
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE (PARC) EN
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV
; TITLE OF INVENTION: THEREMITH
; FILE REFERENCE: N1AD 201
; CURRENT APPLICATION NUMBER: US/09/511,507
; CURRENT FILING DATE: 2000-02-23
; PRIOR APPLICATION NUMBER: 09/302,812
; PRIOR FILING DATE: 1999-04-30
; NUMBER OF SEQ ID NOS: 38
; SEQ ID NO 8
; LENGTH: 768
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
; FEATURE:
US-09-511-507-8

```

[illegible]

```
Db 331 EHLIDMSQAAPLGVPLVHDAGETIEDSGILGQVDFANKVLGGVHGCVQEEIRV 390
QY 762 INPELIVSLFTEVLVDHNECLITGTQEYSEYTYGAYTVMARSHDRSERDWRRTTE 821
Db 391 ICPPELLVGLKFTCLRPPEALVMLGAEYSNTYAGSPWNSGPNFSDTPRDSGRRQTA 450
QY 822 IVIDALHFRYLDQVPFKIRRELKAYCGFLR-----PGVSSENLSAVATGNWCGCA 875
Db 451 IVIDALHFAQSHQVREDLMERELKAVIGFVHMVMTPPPG-----VATGNWCGCA 502
QY 876 FGDGABLKALIQILAAVAERDVVYTFGDSLMRDIYSNHTFLTERKLTGVEYKLLLR 935
Db 503 FGDGSLYKALLQMLVCAQLGRPLAYTTFGNVFRDDFHEWMLLFRNDGTTVQQLWS-ILR 561
QY 936 YNEECRNCSTGP-----DIKLYPIYHVESCTQTNQPCQRTGA 977
Db 562 SYSRLIKESSKEPRENKASKKLYDFI---KEELKKVRDVPQEGASA 606

RESULT 13
US-09-302-812-10
; Sequence 10, Application US/09302812B
; Patent No. 6333148
; GENERAL INFORMATION:
; APPLICANT: JACOBSON, Myron K.
; APPLICANT: JACOBSON, Elaine L.
; APPLICANT: AM, Jean-Christophe
; APPLICANT: LIN, Winston
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE (PARG) ENZ
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIVE
; TITLE OF INVENTION: THEREWITH
; FILE REFERENCE: NIAD 201
; CURRENT APPLICATION NUMBER: US/09/302,812B
; CURRENT FILING DATE: 1999-04-30
; EARLIER APPLICATION NUMBER: 60/083,768
; EARLIER FILING DATE: 1998-05-01
; NUMBER OF SEQ ID NOS: 38
; SEQ ID NO 10
; LENGTH: 726
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
; FEATURE:
US-09-302-812-10

Query Match 9.0%; Score 465; DB 4; Length 726;
Best Local Similarity 24.3%; Pred. No. 3.5e-34;
Matches 204; Conservative 116; Mismatches 282; Indels 236; Gaps 35;

QY 199 DTNSESRRNQOPLTHVKLANAKQTM-----DEQGREARSHQKCGKACHPAEACAGCQOE 254
Db 10 DPVTQDEKDYEDYV-GVGFHQVPTMKRKLTEHGNNTESKE-----DPEE-----PK 56
QY 255 ETDVVSPLSDTGSDE-----VGTGLKNA-----NRLNRQESSLGN 291
Db 57 SRDVFVSSQSSDESQEDSAENPEIAKEVSENCENLTETLKISNIESLDNVTSEHTLDN 116
QY 292 ---SPPEKSESESPMDV-----DNSKSCQSDSEADETSPGF--DEQE--- 331
Db 117 HKSTPEEDVNNKSNIDVAINSDDEDELVLKENNEMRDEGQVQQLSODLFADDQELIE 176
QY 332 -----DSSAOTANKPSFOPREADELTKRSKAGGEBIRLHFQEGGESR 377
Db 177 YPGIMKDTTQDITDSEVETAQOMIEETEADS-----TFVGDSK 219
QY 378 AGMNDVNAKRGSTSSLNVECRNSKHGRKDSKITDHFMRVPAKEDKRXQCEMKKHORTE 437
Db 220 A-----TKTVRTSSSFLST-----VSTCEAPAKGRARMYQKELE 254
QY 438 RKIPKYPHLS--PDKKWLGTPTEEMRMRPCGIRLPPLRPSANHTVTRVLLRIGEV 495
Db 255 KHVIAFTEGNLTQPD-----LNKVDPRNY-----RYCTI 285
```

```
QY 496 KPFPF-THFKDLWDNKH---VKMPCSEQNLYFVEDENGERAAGSRWELIQTALLNLRTRP 551
Db 286 PN-FPASQCKLRDNRYGPKIVLPQRWREF---DSRGR----- 320
QY 552 QNLKDAILKYN---VAYSKWDFTA-----LIDFWD---KVLBEAEAHLYQSILPDM 598
Db 321 ---RDSYFVKRKLGDYLCYKTTGTGFMFVGLLHNNWEEDPDITTYKLPALEMYKEMSEL 377
QY 599 V-----KIALCLPNICTQPIPLAKQKMHSTMSQEQIASLILANAFFCTFPR 645
Db 378 VGREVELEKFAVARIAKTAEDILPERIYRLVDV-ESATLSHKQCAALVARMFF----- 431
QY 646 RNAKMKSEYSSYPDINFNLFEGRSSRKEPKLTLFCYFRRVVEKKPTGLVTF--TRQSL 703
Db 432 --ARPDSPFS-----FCRILSSDKSICVEKLFKLTFTYFDKMSMDPPDGAVSFRLTKMDK 483
QY 704 EDPPEWERCEKLLRLHVTY--EGTIEGNGQGMLOVDFAHRFVGCGVTSAGLVQEBIRFL 761
Db 484 DTFNEEWKDKLRLSPEVEFFDDEMLIEDTAL-CTQVDFANEHLGGGVNLHSGVQEBIRFL 542
QY 762 INPELIVSLFTEVLVDHNECLITGTQEYSEYTYGAYTYRWA-----RSHEDRSE-RDDW 815
Db 543 MCPENVMGMLLCEKMKQLEAISIVGAYVSSYTGHTLKWAELOPNHSRQNTNEFRDRF 602
QY 816 QRRTEIVAIDALHFR-----YLDQFVPEKIRRELKAYCGFLRPGVSSENLSAVATGN 870
Db 603 GLRLVETIAIDAILFKGSKLDCQTEQLNKANITIREMKKASIGFMSQGPXTNI-PIVTGW 661
QY 871 WCGGARGGABRKALIQILAAVAERDVVYTFGDSLMRDIYSNHTFLTERKLTGVE 928
Db 662 WCGGAFNGDKPLFTFIQVIAAGVADRPLHFCFGEPELAACKKIIRMKQKDVTLGK 719

RESULT 14
US-09-511-477-10
; Sequence 10, Application US/09511477
; Patent No. 6337202
; GENERAL INFORMATION:
; APPLICANT: JACOBSON, Myron K.
; APPLICANT: JACOBSON, Elaine L.
; APPLICANT: AME, Jean-Christophe
; APPLICANT: LIN, Winston
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE (PARG) EN
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV
; TITLE OF INVENTION: THEREWITH
; FILE REFERENCE: NIAD 201
; CURRENT APPLICATION NUMBER: US/09/511,477
; CURRENT FILING DATE: 2000-02-23
; PRIOR APPLICATION NUMBER: 09/302,812
; PRIOR FILING DATE: 1999-04-30
; NUMBER OF SEQ ID NOS: 38
; SEQ ID NO 10
; LENGTH: 726
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
; FEATURE:
US-09-511-477-10

Query Match 9.0%; Score 465; DB 4; Length 726;
Best Local Similarity 24.3%; Pred. No. 3.5e-34;
Matches 204; Conservative 116; Mismatches 282; Indels 236; Gaps 35;

QY 199 DTNSESRRNQOPLTHVKLANAKQTM-----DEQGREARSHQKCGKACHPAEACAGCQOE 254
Db 10 DPVTQDEKDYEDYV-GVGFHQVPTMKRKLTEHGNNTESKE-----DPEE-----PK 56
QY 255 ETDVVSPLSDTGSDE-----VGTGLKNA-----NRLNRQESSLGN 291
Db 57 SRDVFVSSQSSDESQEDSAENPEIAKEVSENCENLTETLKISNIESLDNVTSEHTLDN 116
QY 292 ---SPPEKSESESPMDV-----DNSKSCQSDSEADETSPGF--DEQE--- 331
Db 117 HKSTPEEDVNNKSNIDVAINSDDEDELVLKENNEMRDEGQVQQLSODLFADDQELIE 176
```

332 -----DSSAQTANKPSRFPQPREADTELKRSKAGGEIRLHFQEGESR 377  
177 YPGIMKOTTQDITDSEVETAQMEMIEETEADS-----TFVGDSK 219  
378 AGMNDVNAKRPGSTSSLNVECRNSKQHGKDSKITDHFMRVPKABDKRKECEMKHQRTE 437  
220 A-----TKTVRTSSSFLST-----VSTCEAPAKGRARMYQKELE 254  
438 RKIPKYPPIPHLS--PDKKWLGTPIEEMRMRPCGIRLPLRPSANHTVTIRVDLLRIGEV 495  
255 KHVIATFEGNLTQPD-----LNKVDPRDNY-----RYCTI 285  
496 PKPPP-THFKDLWNKH--VKMPCSEQNLYPVEDENGERAAGSRWELIQTALLNRLTRP 551  
286 PN-PPASQGLREDNRYGPKIVLPORWREF-----DSRGR----- 320  
552 QNLKDAILKYN--VAYSCKWDFTA-----LIDFWD-----KVLSEABAQHLYSILPDM 598  
321 ---RDSYFYFKRKLGYLKYKTTGYFMFVGLLHNWFEFDPDITYKLPALEMYIKEMSEL 377  
599 V-----KIALCLPNICTOPTPLKQKMNHSITMSOEOIASILANAFCTPFR 645  
378 VGREVLEKFAVARIAKTAEDILPERIYRLVGDV-ESATLSHKQCAALVARMFP----- 431  
646 RNAMKMSYSSYPDINFNRLPEGRSSRKPDKLTCFYFRRVTEKKPTGLVTF--TROSL 703  
432 --ARPDSPFS-----FCRILSSDKSICVEKLFYFDKMSMDPPDGAVSFRLTMDK 483  
704 EDFPEWERCCELLTRLHVTY--EGTIEGNQGMLOVDPANRFPVGGVTSAGLVQBEIRFL 761  
484 DTFNEWKDKKPLKFLFIQVIAAGVADRPHFCSGPELAACKKIIERMKQKDVILGK 719  
762 INPELIVSRFLTEVLDHNECLITTEQYSEYTGVAETYRWA-----RSHEDRSE-RDDW 815  
543 MCPENVMVGMMLCEKMKQLEAISIVGAYVFSSTGYGHTLKWAELOPNHSRQNTNEFRDRF 602  
816 QRRTEIIVADALHFR-----YLDQFVPEKIRRELKAYCGFLRPGVSSNLSAVATGN 870  
603 GLRLVETIADAILFKGSKLDQCTEQNLKANIIREMKKASIGFMSQGPFTNI-PIVTGW 661  
871 WCGGAFGGDARUKALIQILAAVAERDVVYFTFGSELMRDIYSMHTFLTERKLTVGE 928  
662 WCGGAFNGDKPLKFLFIQVIAAGVADRPHFCSGPELAACKKIIERMKQKDVILGK 719

RESULT 15  
US-09-511-507-10  
; Sequence 10, Application US/09511507  
; Patent No. 6395543  
; GENERAL INFORMATION:  
; APPLICANT: JACOBSON, Myron K.  
; APPLICANT: JACOBSON, Elaine L.  
; APPLICANT: AME, Jean-Christophe  
; APPLICANT: LIN, Winston  
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE (PARG) EN  
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV  
; TITLE OF INVENTION: THEREWITH  
; FILE REFERENCE: NIAD 201  
; CURRENT APPLICATION NUMBER: US/09/511,507  
; CURRENT FILING DATE: 2000-02-23  
; PRIOR APPLICATION NUMBER: 09/302,812  
; PRIOR FILING DATE: 1999-04-30  
; NUMBER OF SEQ ID NOS: 38  
; SEQ ID NO 10  
; LENGTH: 726  
; TYPE: PRT  
; ORGANISM: Caenorhabditis elegans  
; FEATURE:  
US-09-511-507-10

Query Match 9.0%; Score 465; DB 4; Length 726;  
Best Local Similarity 24.3%; Pred. No. 3.5e-34;

Matches 204; Conservative 116; Mismatches 282; Indels 236; Gaps 35;  
199 DTNSEERDNOQFLTHVKLANAKQTM-----DEQREARSHQKCGKACHPAEACAGCQOE 254  
10 DPVTQDEKDYEDYV-GVGFAHQVPTMKRKLTEHNGNTESKE-----DPE-----PK 56  
255 ETDVVSSEPLSDTGSSED-----VGTGLKNA-----NRLNROESSIGN 291  
57 SRDVFVSSQSDSEQSDSAENPEIAKEVSENCENLTETLKISNTESLDNVTERSEHTLDN 116  
292 ---SPPEKSESEBPMNV-----DNSKNSQDSEADEETSPGF--DEQE--- 331  
117 HKSTEPMEEDVNNKSNIDVAINDEDEDELVEENKEMRDGEQVQQLSFADQELIE 176  
332 -----DSSAQTANKPSRFPQPREADTELKRSKAGGEIRLHFQEGESR 377  
177 YPGIMKOTTQDITDSEVETAQMEMIEETEADS-----TFVGDSK 219  
378 AGMNDVNAKRPGSTSSLNVECRNSKQHGKDSKITDHFMRVPKABDKRKECEMKHQRTE 437  
220 A-----TKTVRTSSSFLST-----VSTCEAPAKGRARMYQKELE 254  
438 RKIPKYPPIPHLS--PDKKWLGTPIEEMRMRPCGIRLPLRPSANHTVTIRVDLLRIGEV 495  
255 KHVIATFEGNLTQPD-----LNKVDPRDNY-----RYCTI 285  
496 PKPPP-THFKDLWNKH--VKMPCSEQNLYPVEDENGERAAGSRWELIQTALLNRLTRP 551  
286 PN-PPASQGLREDNRYGPKIVLPORWREF-----DSRGR----- 320  
552 QNLKDAILKYN--VAYSCKWDFTA-----LIDFWD-----KVLSEABAQHLYSILPDM 598  
321 ---RDSYFYFKRKLGYLKYKTTGYFMFVGLLHNWFEFDPDITYKLPALEMYIKEMSEL 377  
599 V-----KIALCLPNICTOPTPLKQKMNHSITMSOEOIASILANAFCTPFR 645  
378 VGREVLEKFAVARIAKTAEDILPERIYRLVGDV-ESATLSHKQCAALVARMFP----- 431  
646 RNAMKMSYSSYPDINFNRLPEGRSSRKPDKLTCFYFRRVTEKKPTGLVTF--TROSL 703  
432 --ARPDSPFS-----FCRILSSDKSICVEKLFYFDKMSMDPPDGAVSFRLTMDK 483  
704 EDFPEWERCCELLTRLHVTY--EGTIEGNQGMLOVDPANRFPVGGVTSAGLVQBEIRFL 761  
484 DTFNEWKDKKPLKFLFIQVIAAGVADRPHFCSGPELAACKKIIERMKQKDVILGK 542  
762 INPELIVSRFLTEVLDHNECLITTEQYSEYTGVAETYRWA-----RSHEDRSE-RDDW 815  
543 MCPENVMVGMMLCEKMKQLEAISIVGAYVFSSTGYGHTLKWAELOPNHSRQNTNEFRDRF 602  
816 QRRTEIIVADALHFR-----YLDQFVPEKIRRELKAYCGFLRPGVSSNLSAVATGN 870  
603 GLRLVETIADAILFKGSKLDQCTEQNLKANIIREMKKASIGFMSQGPFTNI-PIVTGW 661  
871 WCGGAFGGDARUKALIQILAAVAERDVVYFTFGSELMRDIYSMHTFLTERKLTVGE 928  
662 WCGGAFNGDKPLKFLFIQVIAAGVADRPHFCSGPELAACKKIIERMKQKDVILGK 719

Search completed: May 26, 2004, 18:49:20  
Job time : 35.3065 secs

hypothetical prote  
erythrocyte-bindin  
hypothetical prote  
chromogranin B pre  
TCOF1 protein - mo  
profilaggrin - hum  
hypothetical prote  
serine/threonine pr  
hypothetical prote  
hypothetical prote  
myosin-IXB (simila  
touscan gene protei  
ankyrin related pr  
dentin matrix acid  
thyroid hormone re  
hypothetical prote  
probable secreted  
DNA-directed DNA p  
probable myosin he  
hypothetical prote  
hypothetical prote  
nestin - golden ha  
hypothetical prote  
hypothetical prote  
AAS surface protei  
nucleolin homolog  
hypothetical prote  
SRP40 protein - ye  
neurofilament trip  
probable GTPase-ac  
NF-180 - sea lamp

15 170 2.3 1071 2 B85343  
16 168 2.3 778 2 A35970  
17 168 2.3 1115 2 T19137  
18 167.5 2.3 646 1 S15901  
19 165.5 2.3 1320 2 JC5630  
20 165 2.3 2248 2 A35938  
21 162.5 2.2 7160 2 T27935  
22 162 2.2 1231 2 T18532  
23 161.5 2.2 943 2 S54493  
24 161.5 2.2 1535 2 T49042  
25 161 2.2 2022 2 A59256  
26 160 2.2 2176 2 T13806  
27 158.5 2.2 1016 2 T19006  
28 158 2.2 489 2 A45988  
29 158 2.2 1424 2 T03851  
30 157.5 2.2 1791 2 T02345  
31 157 2.2 1192 2 A71623  
32 157 2.2 3122 2 T17202  
33 156.5 2.2 1269 2 F84730  
34 156 2.2 1222 2 T22490  
35 156 2.2 1280 2 T00365  
36 156 2.2 1804 2 T34518  
37 156 2.2 3147 2 T18674  
38 155.5 2.1 522 2 C96608  
39 155.5 2.1 1463 2 T30290  
40 154.5 2.1 611 2 T06458  
41 154.5 2.1 1877 2 T21861  
42 154 2.1 406 2 S38170  
43 154 2.1 854 2 S02003  
44 154 2.1 1822 2 T14106  
45 153.5 2.1 1110 2 I51116

# ALIGNMENTS

## RESULT 1

B84726  
probable poly(ADP-ribose) glycohydrolase [imported] - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cross)  
C>Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 02-Feb-2001  
C:Accession: B84726  
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;  
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.;  
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.  
Nature 402, 761-768, 1999  
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.  
A:Reference number: A84420; MUID:20083487; PMID:10617197  
A:Accession: B84726  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-997 <STO>  
A:Cross-references: GB:AE002093; NID:g4887749; PIDN:AAD32285.1; GSPDB:GN00139  
C:Genetics:  
A:Gene: At2g31870  
A:Map position: 2

Alignment Scores: 6.62e-28 Length: 997  
Pred. No.: 522.00 Matches: 136  
Score: 45.88% Conservative: 59  
Percent Similarity: 32.00% Mismatches: 116  
Best Local Similarity: 7.21% Indels: 114  
Query Match: 2 Gaps: 12  
DB:

US-09-302-812-1 (1-4070) x B84726 (1-997)

QY 1989 TGGGATAAGTACTACAGAGCAGAGCTCAACATTGTATCAGTCCATCTTCCTGAT 2048  
Db 79 PheAspGluLeuIleAspGluLysGluSerLysArgTrpPheAspGluIleIleProAla 98  
QY 2049 ATGGTGAATTCGATCTCTCTGCCAATATT----- 2081  
Db 99 LeuAlaSerLeuLeuGluGlnPheProSerLeuGluValHisPheGlnAsnAlaAsp 118

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: May 26, 2004, 16:16:17 ; Search time 59.0768 Seconds  
(without alignments)  
13253.917 Million cell updates/sec

Title: US-09-302-812-1

Perfect score: 7242

Sequence: 1 accggaagtagaacgaccc.....aaatttcattcaacaaaaa 4070

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 283366 segs, 96191526 residues

Total number of hits satisfying chosen parameters: 566732

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ n2p.model -DEV=xlp  
-Q=/cgn2.1/USPTO spool\_p/6331148/runat 26052004 150054 5689/app query.fasta\_1.12437  
-DB=PIR 78 -QFWT=fastan -SUFFIX=n2p.rpr -MINMATCH=0.1 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=bloum62 -TRANS=human40.cdi -LIST=45  
-DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=6331148 @CGN 1 1 298 @runat 26052004 150054 5689 -NCPU=3  
-NO WMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :  
1: pir1.\*  
2: pir2.\*  
3: pir3.\*  
4: pir4.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	522	7.2	997	B84726	probable poly(ADP-
2	465	6.4	726	T21138	hypothetical prote
3	331	4.6	364	A84726	probable poly(ADP-
4	185.5	2.6	1320	S57113	BUD4 protein - yea
5	179	2.5	1213	A58198	serine/proline-ric
6	175.5	2.4	3942	T42730	Bassoon protein -
7	175.5	2.4	5327	T13564	microtubule-associ
8	175	2.4	1359	T34036	hypothetical prote
9	174.5	2.4	1403	A47328	natural killer cel
10	173	2.4	1210	I39410	AP-4 protein, spli
11	170.5	2.4	1200	A46194	neurofilament prot
12	170.5	2.4	1507	B47328	natural killer cel
13	170.5	2.4	1576	T03277	pol protein - yeas
14	170.5	2.4	6642	T29757	protein UNC-89 - C







A;Accession: A84726  
 A;Status: preliminary  
 A;Molecule type: DNA  
 A;Residues: 1-364 <STO>  
 A;Cross-references: GB:AE002093; NID:g4887750; PIDN:AAD32286.1; GSPDB:GN00139  
 C;Genetics:  
 A;Gene: AC2g31860  
 A;Map position: 2

Alignment Scores:  
 Pred. No.: 8.04e-15 Length: 364  
 Score: 331.00 Matches: 110  
 Percent Similarity: 39.77% Conservative: 63  
 Best Local Similarity: 25.29% Mismatches: 140  
 Query Match: 4.57% Indels: 122  
 DB: 2 Gaps: 12

US-09-302-812-1 (1-4070) x A84726 (1-364)  
 QY 2001 CTGAGAGAGCAGAGAGCTCAACACTTGTATCATGCTCCATCTTGGCTGATATGGTGAATAATT 2060  
 Db 1 MetSerLysGlnGluSerArgTrpPheAsnGluPheLeuProAlaMetAlaCysLeu 20  
 QY 2061 GCACCTGTCTGCCAATATTGTGACCCAG----- 2090  
 Db 21 LeuLeuArgPheProSerLeuLeuGluSerHisTyrLeuAsnSerAspAsnLeuLeuAsn 40  
 QY 2091 -----CCAAATACCACTCTCAACAGAGATGATCATCTCCATCAACAATGCA 2138  
 Db 41 GlyThrLysThrGlyLeuArgValLeuValProAsnLysAlaGlyIleValPheLeuSer 60  
 QY 2139 CAGAAACAGATTGCGAGCTTTAGCTAAATGCTTTCTTCTGCGAGTTTCCACGACGCAAT 2198  
 Db 61 GlnGluLeuIleGlyAlaLeuLeuSerCysSerPhePheCysLeuPhePro----- 77  
 QY 2199 GCCAAGATGAATCAGAGTATTCAGTTAT---CCAGATATTACTCAATCGGTGTTT 2255  
 Db 78 -----ValAspAspArgGlySerAsnHisLeuProIleIleAsnPheAspLysLeuPhe 95  
 QY 2256 -----GAAGGAGCTTCATCAAGGAAACACAGAGAGCTTAAACCGCTCTTC 2300  
 Db 96 GlySerLeuLeuAsnThrGlyArgAsnGluHisGlnGluAsnLysIleLysCysIleIle 115  
 QY 2301 TGCTACTTTAGAGAGTCACAGAAACCCACTGGTGTGGTGCACATTCACAAGACAG 2360  
 Db 116 HisTyrPheGlnArgLeuSerSerSerIleSerProGlyPheValSerPheGluArgLys 135  
 QY 2361 -----AGTCTTGAA-----GATTTCCAGAGTGGGAAAGATGTGAA 2396  
 Db 136 IleLeuSerLeuGluGlnAspSerSerThrLeuAspGluGlyPheTrpGlyLysSerThr 155  
 QY 2397 AAACCTCGTACTCGACTGCATGTCTACATTACAGGTACCATAGAGAAACGGCCAGGC 2456  
 Db 156 ValAsnLeuCysProValGluValArgThrSerGlyLeuIleGluAspGlnSerValGlu 175  
 QY 2457 ATGCTACAGGTGGATTTTGCACACCGTTTCTGAGGTGGTGTGTAACGAGTCGAGACTT 2516  
 Db 176 AlaLeuLeuValAspPheAlaAsnLysAsnLeuGlyGlyAlaLeuArgLysGlyCys 195  
 QY 2517 GTGCAAGAGAAATCCGCTTTTAAATCAACCTGAGTTGATGTTTCACGGCTCTTCACT 2576  
 Db 196 ValGlnGluGluIleArgPheMetIleAsnProGluLeuIleValGlyMetPheLeu 215  
 QY 2577 GAGGTGCTGGATCAAAATGATGCTTATCATCAGGTACTGACAGTACAGTGAATATAC 2636  
 Db 216 ProThrMetGluValThrGluAlaIleGluValGlyAlaGluArgPheSerLeuTyr 235  
 QY 2637 ACAGCTATGCCGAAACATACCGCTGGGCCGAGCCATGAAGACAGGAGCGAAAGGAC 2696  
 Db 236 ThrGlyCys----- 238  
 QY 2697 GACTGGCAGAGCGCAGCAGTGAATCGTTCATTCGACGCGCTCCACTTCAGACGCTAC 2756  
 Db -----

Db 239 ---PheArgLysAlaLysThrArgIleValAlaIleAspAlaLeuArg----- 253  
 QY 2757 CTCGACAGCTTTGTGCCGAGAGATCAGACGGGAGCTTAAACAAGCTTACTGTGGATTT 2816  
 Db 253 ----- 253  
 QY 2817 CTTTCGTCTGGAGCTTTCTTCAGAGAACCTGTCTGCGAGTGGCTACAGAAACTGGGCTGT 2876  
 Db 254 ---HisProGlyValSer----- 258  
 QY 2877 GGTGCTTTGGGGGTGATGCTAGACTAAAGCCTTAATACAGATCTCGGAGCTGCTGTA 2936  
 Db 259 -----GlnTyrLysLeuGluSerLeuLeuSerValLeuIleLeuSerSer 273  
 QY 2937 GCTGAGCGAGACGTGGTGTATTATTTTACC-----TTTGGGACTCA 2975  
 Db 274 SerGlyArgProIleArgLeuTyrMetGlySerValSerLeuGlnGlyIleGlyAspVal 293  
 QY 2976 GAACCTGATGAGACATTTTACAGCATGCATACATTTCTCTACTGAGAGAACTGACTGTT 3035  
 Db 294 ValLeuMetValGluIleLeuSerSerLeuPhePheAsnGlyLeuArgPheHisArg 313  
 QY 3036 GGAGAGTATATAAGCTG----- 3053  
 Db 314 SerAsnLeuTyrLeuPheSerProLeuIleAlaCysAsnTyrGlnProMetValValVal 333  
 QY 3054 ---CTGCTACGATATTACATGAAAGATCGAAGAACTGCTCCACCCCGGACGACATC 3110  
 Db 334 TrpLeuThrThrLeuHisSerPheGlnAlaArgArg-----ProPheMet 348  
 QY 3111 AAGCTTATTCATTCATATACATGATGAGTGTGATGCTCTGTCACACAG 3155  
 Db 349 SerTyrTyrThrPheGlyPheGluAlaLeuGlnAsnLeuAsnGln 363

RESULT 4  
 S57113  
 BUD4 protein - yeast (Saccharomyces cerevisiae)  
 N;Alternate names: protein J1905; protein YUR092w  
 C;Species: Saccharomyces cerevisiae  
 C;Date: 08-Jul-1995 #sequence\_revision 08-Sep-1995 #text\_change 29-Oct-1999  
 C;Accession: S57113; S61952  
 R;Ramezani Rad, M.; Kirchrath, L.; Hollenberg, C.P.  
 submitted to the Protein Sequence Database, September 1995  
 A;Reference number: S57111  
 A;Accession: S57113  
 A;Molecule type: DNA  
 A;Residues: 1-1320 <RAM>  
 A;Cross-references: EMBL:Z49592; MIPS:YUR092w  
 R;Sanders, S.L.; Herzkowitz, I.  
 submitted to the EMBL Data Library, November 1995  
 A;Reference number: S61952  
 A;Accession: S61952  
 A;Molecule type: DNA  
 A;Residues: 'MHAESTVDLKKEDNEMETKSNITQNGSEDTFHNWKLPLQETGDDTMEMLVKHNTSRNATENSRCRSP', S  
 202-212, 'D', 214-1007, 'A', 1009-1320 <SAN>  
 A;Cross-references: EMBL:U41641; NID:g1151116; PIDN:AAE17116.1; PID:g1151117  
 C;Genetics:  
 A;Gene: SGD:BUD4  
 A;Cross-references: MIPS:YUR092w; SGD:S0003852  
 A;Map position: 10R

Alignment Scores:  
 Pred. No.: 0.000122 Length: 1320  
 Score: 185.50 Matches: 185  
 Percent Similarity: 31.20% Conservative: 156  
 Best Local Similarity: 16.93% Mismatches: 361  
 Query Match: 2.56% Indels: 391  
 DB: 2 Gaps: 45

US-09-302-812-1 (1-4070) x S57113 (1-1320)  
 QY 465 GCCACCTCTCTGTTTTCACACAGAGACTATACACAGTTGGATGGACACTAAA----- 518  
 Db -----

Db SerThrSerThrLysSerGluGluSerTyrIleAlaAspTyrLysValThrArgGlnGlu 309  
QY 519 -----GGAATCAAGACAGCTGAATCAGAAAGTTTGCATAGTAAAGAAAC----- 563  
Db AspTpaAspThrLysLysLeuHisGlnGluSerGluHisAlaAsnGluGlnProAlaIle 329  
QY 564 -----AACAAACAGAGAGAAATCCATGATGAGTTCGTGTACAAAAGATTAACCTTTAT 617  
Db IleProGlnLysAspSerSerGluGluThrPheThrGluLeuAsnGluSerGluPhe 349  
QY 618 CACATAACATG----- 629  
Db GlnArgAsnPheLysAspGlyGluGluTyrArgIleValGlnHisGluGluSerLeuTyr 369  
QY 630 -----GAAAAATAGAAAATGTTTCTCAGCTAGTGTTTGAT 665  
Db GlyGlnArgThrLysSerProGluGluAsnIleAlaAsnGlySerGluIleGlyValAsp 389  
QY 666 -----AAGTCACCATGTTGAAAAGGTACACAGTATTGAAAGCAG 704  
Db HisGlyGluAlaAlaGluValAsnGluProLeuAlaLysThrSer-----AlaGluGlu 407  
QY 705 CATCAGACTCGCGCTATGTAGTGCAG-----AATGAAGGGCCACACTCAGAACG3 758  
Db HisAspLeuSerSerCysGluAspGlnSerValSerGluAlaArgAsnLysAspArg 427  
QY 759 CTTTTCGGAA----- 767  
Db IleGluGluLysGluValGluThrLysAspGluAsnIleGluThrGluLysAspGluSer 447  
QY 768 -----ATGAACCTCCAGCGGTAACTCTGTGTACAGAG----- 800  
Db GluTyrHisLysValGluGluAsnGluProGluHisValProLeuLeuProLeu 467  
QY 801 -----CAGTTCAGTAATGCTGATCGATCAGTCAGTCGCTCCCAAGGAT 842  
Db ProArgTyrGluGluIleGlnPheAsnGluProPheIleAspGluAsnAspThrSerAsn 487  
QY 843 GATCAGATGAC----- 854  
Db AspSerIleAspLeuThrArgSerMetLysProSerAspTyrIleSerIleTyrHisIle 507  
QY 855 -----ACAATAGTAGGAGAGTAGAGTAATCAGCAGTTTGTACA 896  
Db GlnGluGluGluIleLysSerAsnSerProGluSerIleAlaAsnSerGlnPheSerGln 527  
QY 897 CATGTAAGCTTCGGAATGCAAGCAGCATGGAAGATGAACAGGCGCAGAGAGCCAGA 956  
Db GlnSerSerIleThrThrAla-----SerThrValAspSerLysAspAsnGlySerThr 546  
QY 957 AGCCACCAG-----AAGTGTGGCAAGCTTGCATCTCTGCAGAACGCC 998  
Db SerPheLysPheLysProArgIleValSerArgSerArgIleTyrAsnProLysSerArg 566  
QY 999 TGTGCAAGGTGTGAG-----CAGGAGGACAGACGTGTGTCCGAGAGCCCC----- 1046  
Db ValSerSerLeuAsnTyrTyrAspAsnGluAspTyrIleLeuSerAsnSerGluTyrAsn 586  
QY 1047 -----TTGTCGACACTGGCTCTCAGGATGTGTGT 1076  
Db AlaLeuAspProMetArgAsnThrLeuIleSerLysArgIleGlnAspAsnIleArg 606  
QY 1077 ACTGGACTGAAAATGCC-----AACAGATTGATAGACAGAA 1115  
Db ThrGlnLysGlyHisAlaProLeuIleArgProSerIleMetLysLeuAsnGlyGluAsp 626  
QY 1116 AGTAGTCTAGGAATCTCTCCATTGTGAGAAAGATGAACCTCAGTCACCAATGGAT 1175  
Db SerGlyPheGlnAsn-----HisPheLeuGluValGluGlnProGlnGlu 641  
QY 1176 GTAGATAATTCCAAAAATGTTCTCAGGATTGAGAGCAGATGAGAGACAGTCCAGGT 1235  
Db HisGluAsnIleProLeuSerThrHisLeuSerGluGlnAspIleThrAsnValGly 661

QY 1236 TTTGATGAACAGGAAGATAGCAGTTCTGCTCAACACAGCAATAAACCTTCAAGGTTCCAA 1295  
Db LeuAspGluGlnLeuProThrAsnThrGlnAsp----- 673  
QY 1296 CCAAGAGAGCTGACACTGAGTTGAGGAGCGGTCTCTGCT----- 1337  
Db GluAlaGluIleSerIleArgGluIleGluSerAlaGlyAspIleThrPheAsn 691  
QY 1338 AAGGAGGTGAGATTTCGATTTCATTTTCAATTTGAAGGAGGAGAGTGCAGTCGGAATG 1397  
Db ArgGlyAspLeuLeuSerLeuSerPheAspGluGluLeuGlnAspPheAlaAsnPhe 711  
QY 1398 AATGATGTG-----AAT 1409  
Db LeuAspAlaLeuAspHisAspSerThrSerPheAsnHisGlyProAspAspSerSerSer 731  
QY 1410 GCCAAACACCTGGAAGTACTTCTAGCCTGAATGTAGAGTCAGCAAAATCTAAGCAA----- 1466  
Db PheGlnArgAspSerSerLysLysSerPheAsnSerLeuTyrGluSerSerTyrGluLeu 751  
QY 1467 -----CATGGGAGAAAGGATTCTAAATTCACAGATCATTTCATGAGAGTCC 1514  
Db LysProProSerIleArgLysGlnProIleAlaProAspValLeuGlnLysLeu----- 770  
QY 1515 AAGCAGAGGAGCAAAAGAAAGCAATGTGAATGAACATCAAGACAGAGAGAGG----- 1571  
Db LeuGluSerAspThrLysAspAlaAspLeuGluLysIleArgGluGluArgIle 789  
QY 1572 -----AAGATCCCT----- 1580  
Db ThrGluProArgThrGlyLeuGlyIleGlyMetLeuLysThrProValLysAspValSer 809  
QY 1580 ----- 1580  
Db IleAlaLeuAlaSerIleLysGlyTyrGluAlaSerPheSerAspThrAspSerArg 829  
QY 1580 ----- 1580  
Db ProGluGlyMetAsnAsnSerAspAlaIleThrLeuAsnMetPheAspAspPheGluGlu 849  
QY 1581 -----AAATACATTCACCT-----CACCTTTCAGATAAGAAATGGCTTGA 1625  
Db AspLysMetThrProSerThrProValArgSerIleSerProIleLysArgHisValSer 869  
QY 1626 ACTCCT----- 1631  
Db SerProPheLysValLysAlaGlyAsnLysGlnGluAsnAsnGluIleAsnIleLys 889  
QY 1632 -----ATTGAGAGATGAGGAGAAATGCCAGGTGTGGATCCGG-----CTG 1673  
Db AlaGluGluGluIleGluProMetThrGlnGlnGluThrAspGlyLysGlnAspIle 909  
QY 1674 CTTCCCTTCAGACCATCTCCCAATCACACAGTACTATTTCGGGTAGATCTTTTGGCAATA 1733  
Db ProProLeuLeuAlaGlnThrLysAspAsnValGluAlaLysGluGluThrIleThrGln 929  
QY 1734 GGAGAGTTCCTAAACCTTTCCCAACACATTTTAAAGAT----- 1772  
Db LeuGluGluProGlnAspValGluGlnGluPheProAspMetGlyThrLeuTyrLeuSer 949  
QY 1773 -----TTGTGGGACACACATGTTAAGATGCTTTCCTTCA 1808  
Db IleLysAlaIleSerThrLeuAlaLeuTyrGlyThrLys-----Ser 963  
QY 1809 GAACAAACTTGTACCTCTGGAAGATGAGATGTGTAGAGTGCAGCGAGCCGCGTGG 1868  
Db HisArgAlaThrTyrAlaIleValPheAspAsnGlyGluAsnValGlnThrProTyr 983  
QY 1869 GAACCTCATT----- 1877  
Db GluSerLeuProTyrAspGlyAsnIleArgIleAsnLysGluPheGluLeuProIleAsp 1003

QY 1978 -----CAGACTGCACCTTCTCAACAGCGCTCACTCGGCCCCAGAACCTGAAGATGCT 1928  
 Db 1004 PheLysGlyLysThrGluThrSerSerAlaSerGluArgAspSerLysLysCys 1023  
 QY 1929 ATTCTGAAGTACAATGTGGCATATTCCTAAG---AAATGGGACTTTACAGCTTTGATGAT 1985  
 Db 1024 ValIleThrLeuLysCysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 1043  
 QY 1986 -----TTCTGGGATGAAGGTA 2000  
 Db 1044 LysValProValGlyLysSerPhePheGlyLysThrLysLysLysLysLysLysLysLys 1063  
 QY 2001 CTAGAAGACAGAGAGCTCAACACTTGTATCAGTCCATCTTGCTGATATGGTGAAT 2060  
 Db 1064 ValGlnLysLysProLysGlnAspGluThrAspTyrLeuPheAlaGlnAspGlySerPhe 1083  
 QY 2061 GCACCTCTGTCTGCCAATATTTGTACCCAGCCATACCACTCTCTGAAACAGAGATGAAT 2120  
 Db 1084 AlaArgCys----- 1086  
 QY 2121 CATTCATCAACATGTACAGGAACAGATTGCCAGTCTTTAGCTAATGCTTTCTTCGTC 2180  
 Db 1087 ---GluIleGluIleAsnGluGlu-----PheLeuLysAsnValAlaPheAsn 1101  
 QY 2181 ACGTTTCCAGACCCATGCCAATGCAATCAATCAGATGATTCAGTATTCAGAT----- 2234  
 Db 1102 Thr---SerHisMetHisTyrAsnMetIleAsnLysTrpSerArgIleAlaAspLysIle 1120  
 QY 2235 ATTAACCTCAATCGTTGTTTGAAGGAGCTTCATCAAGAAACAGAGAGCTTAAACG 2294  
 Db 1121 HisGlySerLysArgLysLysLysLysLysLysLysLysLysLysLysLysLysLys 1139  
 QY 2295 CTC-----TTCTGCTACTTTAGAGAGTCTACAGAGAAAAACCCACTGGTGGT 2345  
 Db 1140 LeuAspValGluAlaCysPheLeuGluArgThrSer----- 1151  
 QY 2346 ACATTCACAGACAGACTCTTGAAGATTTCAC---GAGTGGGAAAGATGTGAATACTC 2402  
 Db 1152 -----AlaPheGluGlnPheProLysGlnPheSerLeuValAsnLysIle 1166  
 QY 2403 CTGACTCGACTG-----CATGTCACTTACGAAGGTACCATGAAGAAACGCGC 2450  
 Db 1167 ValSerLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 1186  
 QY 2451 CAGGCGATCTACAGTGGATTTTCACAAACCGTTTCGTT-----GGAGGTGGTGTA 2501  
 Db 1187 ---GlyAspLeuLysGlyLysLysLysLysLysLysLysLysLysLysLysLysLys 1205  
 QY 2502 ACCAGTCGAGCACTTGTGCAGAGAAATCGCTTTTAAATCAACCTGAGTTCATGTT 2561  
 Db 1206 SerGlyTyrHisGluIleSerArgLysAlaLysIleAspIleAsnLeuLysValThr 1225  
 QY 2562 -----TCACGCGCTCTTCACTGAG 2579  
 Db 1226 LysValLeuArgAsnGluAspIleGlnAlaAspAsnGlyGlyGlnArgAsnPheThrAsp 1245  
 QY 2580 GTGCTGATCAATGATGCTTATCATCAAGTACTGACAGCTACAGTGAATACACA 2639  
 Db 1246 TrpValLeuPheAsnGluCysPheGlnLeuValPheAspAsp----- 1259  
 QY 2640 GGCTATGCCGAACATACCGCTGGCGCCGAGCCATGAAGACAGAGCGAAAGGAGCGAC 2699  
 Db 1260 GlyGluArgIleThrPheAsnAlaGluCysSerAsnGluGluLysSer-----Asp 1276  
 QY 2700 TGGCAGAGCGCAGCACTGAGATCGTCGCCATCGACGCGCTCCAC-----TTCAGA 2750  
 Db 1277 TrpTyrAsnLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 1296  
 QY 2751 CGCTACCTCGACCATGTTGTGCGCCGAGAGAGATCAGACGG 2789  
 Db 1297 LysTyrCysGluLysLeuAlaGluGluLysThrArg 1309

RESULT 5

A58198  
 Serine/proline-rich FEL protein, splice form 1 - human  
 N/Alternate names: AF-4 protein  
 C/Species: Homo sapiens (man)  
 C/Date: 02-Jul-1996 #sequence\_revision 25-Oct-1996 #text\_change 01-Dec-2000  
 C/Accession: A58198; 152572  
 R/Morrissey, J.J.M.  
 Blood 81, 1124-1131, 1993  
 A/Title: A serine/proline-rich protein is fused to HRX in t(4;11) acute leukemias.  
 A/Reference number: 152572; PMID:93184301; PMID:8443374  
 A/Accession: A58198  
 A/Status: translated from GB/EMBL/DBJ  
 A/Molecule type: mRNA  
 A/Residues: 1-1213 <MOR>  
 A/Cross-references: GB:L25050; NID:G407323  
 A/Accession: 152572  
 A/Status: translated from GB/EMBL/DBJ  
 A/Molecule type: mRNA  
 A/Residues: 'MAAOS', 13-1213 <MO2>  
 A/Cross-references: GB:L25050; NID:G407323; PIDN:AAA36642.1; PID:G407324  
 C/Comment: This protein is one of several to form chimeric fusion proteins with the trit  
 C/Genetics:  
 A/Gene: GDB:MLLT2; AF-4; FEL  
 A/Cross-references: GDB:136792; OMIM:159557  
 A/Map position: 4q21-4q21  
 C/Keywords: alternative splicing

Alignment Scores:  
 Pred. No.: 0.000337 Length: 1213  
 Score: 179.00 Matches: 238  
 Percent Similarity: 32.72% Conservative: 137  
 Best Local Similarity: 20.77% Mismatches: 449  
 Query Match: 2.47% Indels: 322  
 DB: 57

US-09-302-812-1 (1-4070) x A58198 (1-1213)

QY 79 CGTCTGATAGGCGCTGTTCCGGAGGCTCTCAGACGAGGCTGCAGAGCAGTCAAGCAGCAGC 138  
 Db 170 ArgLeuGlyGlnGluGlyPheGly-----SerSerHisLysLysGlyAspArgArg 187  
 QY 139 GCAGCAGGGGCATGTGTCGCGG-----GAGGCACCGAGGAGGGGGCGGCGGCGCG 186  
 Db 188 AlaAspGlyAspHisCysAlaSerValThrAspSerAlaProGluArgGluLeuSerPro 207  
 QY 187 -----TCCCTCCCGAGGTTAGTG-----AATGAGGCTCTA 216  
 Db 208 LeuIleSerLeuProSerProValProLeuSerProIleHisSerAsnGlnThr 227  
 QY 217 CGCCCGGCTGGCCCGGAGACTCAGTCTCGGGTCCCGAGTCCAGATGAGTGGCG----- 267  
 Db 228 LeuProArgThrGlnGlySerSerLysValHisGlySerSerAsnAsnSerLysGlyTyr 247  
 QY 268 GCCCGGCTGTGAGCCCT-----GCACCAAGC 294  
 Db 248 CysProAlaLysSerProLysAspLeuAlaValLysValHisAspLysGluThrProGln 267  
 QY 295 GAC-----CCCGTGGAGCCCGCTGCACTTCTCCCGCGCGCGCTCGGACG 342  
 Db 268 AspSerLeuValAlaProAlaGlnProProSerGlnThrPhe---ProProProSerLeu 286  
 QY 343 CCGGAGCTTCCCGGCGGAGCGGCTCTCGATTCCAGAGCGC-----TCCGG 396  
 Db 287 ProSerLysSerValAla-MetGlnGlnLysProThrAlaTyrValArgProMetAspGln 306  
 QY 397 TGCAGTTTCAGGTCGCCCGCTCTCTCGTCAGGCTCGCCCTCGCGCGCGCGCGG----- 448  
 Db 306 YGlnAspGlnAlaProSerGluSerProGluLeuLysProLeuProGluAspTyrArgGln 326  
 QY 449 -----ACAGCAGCGAGCGCGCCACTCTCTTGTTCACAGACACTATACAGGTT 504  
 Db 326 nGlnThrPheGluLys-ThrAspLeuLysValProAlaLysAlaLysLeuThrLysLeu 346

QY 505 GGATGCACACTAAAGATCAACAGACAGTTGAATCAGAAAGTTTGCATAGTAAGAA----- 560  
Db ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :  
346 ysMetProSerGlnSerValGluGlnThrTyrSerAsnGluValHisCysValGluGluI 366  
QY 561 -----AACCAACAATA 570  
Db ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :  
366 leLeuLysGluMetThrHisSerTTPProProLeuThrAlaIleHisThrProSerT 386  
QY 571 CAAGAGAGAATCCATGATGAGTCTGTACAAAAAGATAACTTTTATCAACATAACATGG 630  
Db ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :  
386 hrAlaGluProSerLysPheProPheThrLysAspSer-----GlnHisValSers 404  
QY 631 AAAAATTAGAAAATGTTCTCAGCTAGGTTTGTATTAAGTCAACAGTGTGAAAAAGGTACAC 690  
Db ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :  
404 erValThrGlnAsnGlnLysGln-----TyrAspThrSer-----SerLysThrH 419  
QY 691 AGTATTTGAAGCAGCATCAGACTGGCGCTATGTGAAG-----TGCGAGA 735  
Db ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :  
419 isSerAsnSerGlnGlnGlyThrSerSerMetLeuGluAspLeuGlnLeuSerAsps 439  
QY 736 ATGAAGGGCCACACTCAGAACCGCTTTTGGAA-----AGTGAACCTCCA- 779  
Db ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :  
439 erGluAspSerAspSerGluGlnThrProGluLysProSerSerSerAlaProProS 459  
QY 780 --GCGGTAACTCTGTTACAGACGAGTTCAGTAATGCTAAT----- 818  
Db ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :  
459 erAlaProGlnSerLeuProGluProValAlaSerAlaHisSerSerAlaGluSerG 479  
QY 819 -----GTGCATCAGTCTGCCCAAGGATGATCAGTGCACACAAAATAGTAGGAGA 870  
Db ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :  
479 luSerThrSerAspSerAspSerSerAspSerGluSerGluSerSerSerAsps 499  
QY 871 GTAGAGATATCAGCAGTGTTCACACATGTAAAGCTTCGGAATGCAAGACGACGATCG 930  
Db ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :  
499 erGluGluAsnGlu-----ProLeuGluThrProAlaProGluProG 513  
QY 931 AAGATGAACAGGCGAGAGCAGAGCCAGAGGCTGCGAAGGTGCGAAGCTTGCATCTCG 990  
Db ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :  
513 luProProThrThrAsnLysTrpGlnLeuAspAsnTrpLeuThrLysValSerGlnProA 533  
QY 991 CAGAAAGCTGTGAGGCTGTCCAGGAGGAGACAGCTGTGTCCGAGAGCCCTTGT 1050  
Db ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :  
533 laAlaProProGluGly-----ProArgSerThrGluProProArgArgHisProGluS 551  
QY 1051 CGGACACTGGCTCTGAGGATGTGCTACTGCACTGAAATGCCACACATTCATGATGAC 1110  
Db ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :  
551 er---LysGlySerSerAspSerAlaThrSer-----G 561  
QY 1111 AAGAAAGTAGTCTAGCAAAATCTCCATTTGAGAAAGAAAGTGAA---CCTGAGTCAC 1167  
Db ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :  
561 lnGluHisSerGluSerLysAspProProProLysSerSerSerLysAlaProArgAlaP 581  
QY 1168 CAATGATGTAGATAAT---TCCAAAAATAGTTGTGAGATTCAGAGACGATGAAAGAGA 1224  
Db ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :  
581 roProGluAlaProHisProGlyLysArgSerCysGlnLysSerProAla----- 597  
QY 1225 CAAGTCCAGGTTTGTATGAACAGGAGATAGCAGTCTGCTCAACAGCAATAAACCTT 1284  
Db ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :  
598 -----GlnGlnGluProProGlnArgGlnThrValGlyThrLysG 611  
QY 1285 CAAGGTTCCAAACAGAGAGCTGACACTGAGTTGAGGAGCGGTCTCTGCTAAGGGAG 1344  
Db ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :  
611 lnProLysLysProValLysAlaSerAlaArgAlaGlySerArgThrSerLeuGlnGlyG 631  
QY 1345 GTGAGATTCGATTACATTTCCAAATTTGAGGAGAGAGAGTCCAGCTGGATGAATGATG 1404  
Db ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :  
631 lyArg-----GluProGlyLeuLeuProTyrGlySerArgAspG 644  
QY 1405 TGAATGCCAAA---CGACCTTGAAGTACTTCTAGCCTGAATGTAGATGTCAGAAATCTTA 1461  
Db ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :  
644 lnThrSerLysAspLysPro-----LysValL 653  
QY 1462 AGCAACATGGGAGA---AAGGATTCTAAAATCACAGATCATTTTCATGAGAGTCCCAAG 1518

Db ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :  
653 ysThrLysGlyArgProArgAlaAlaAlaSerAsnGluProLysProAlaValProProS 673  
QY 1519 CAGAGGACAAAGAAAGAAACAATGT----- 1544  
Db ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :  
673 erSerGluLysLysLysHisLysSerSerLeuProAlaProSerLysAlaLeuSerGlyP 693  
QY 1545 --GAATGAACATCAAGAACAGAAAGAGATCCCTAAATACATT----- 1589  
Db ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :  
693 roGluProAlaLysAspAsnValGluAspArgThrProGlnHisPheAlaLeuValProL 713  
QY 1590 -----CCACCTCACCTTTCTCCAGATAAGAAATGGCTTGA----- 1625  
Db ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :  
713 euThrGluSerGlnGlyProProHisSerGlySerGlySerArgThrSerGlyCysArgG 733  
QY 1626 -----ACTCCTATTGAGGAGATGAGGAATGCCAAGGTGTGGATC----- 1667  
Db ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :  
733 lnAlaValValGlnGluAspSerArgLysAspLeuProLeuProLeuArgAspT 753  
QY 1668 -----CGGCTGCTCCCTTGAGA-----CCATCTGCCAATCACACAGTGACTATTC 1713  
Db ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :  
753 hrLysLeuLeuSerProLeuArgAspThrProProProGlnSerLeuMetValLysIleT 773  
QY 1714 GGGTAGATCTTTTGGCAATAGCAGAAAGTTCTTAAACCTTTCCCAACACATTTTAAAGATT 1773  
Db ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :  
773 hrLeuAspLeuLeu-----SerArgIleProGlnPro----- 783  
QY 1774 TGTGGGACACAGCATGTTAAGATGCCTTGTTCAGACAAACCTTGATACCTGTG----- 1829  
Db ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :  
784 -----ProGlyLysGlySerArgGlnArgLysAlaGluAspLysGlnProAlaGlyL 802  
QY 1830 -----GAAGATGAGAAATGGTGCAGCAGCTGCAGGCGAGCC 1863  
Db ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :  
802 ysLysHisSerSerGlnLysArgSerSerSerSerSerSerLysLysLeuAlaLysArgL 822  
QY 1864 GTTGGAACTCATTGACACTGCATTCTCAACAGGCTACTCGGCCCCAGACCTGAAGG 1923  
Db ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :  
822 ysGlyGluAlaGluArgAspCysAspAsnLysLysIleArgLeuGluLysGluIleLysS 842  
QY 1924 ATGCTATTCTGAAGTACAAATGGGCATAT----- 1952  
Db ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :  
842 erGlnSerSerSerSerSerSerHisLysGluSerSerLysThrLysProSerArgP 862  
QY 1953 -----CTAAGAAATGGGACTTTACAGCTTTGATTGATTTCTGGGATAAG 1998  
Db ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :  
862 roSerSerGlnSerSerLysLys-----GluM 871  
QY 1999 TACTAGAAAGACAGAGCTCAACTGTATCATCTTGTATCATCTTGTCTGTATGTGTAATA 2058  
Db ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :  
871 etLeuProProProValSerSerSerSerSerSerSerSerSerSerSerSerSerSer 891  
QY 2059 TTGCACCTGTCTGCCAATATTTTACCACCAATACCCTCTCTGAAA-----CAGA 2112  
Db ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :  
891 rgSerArgGluAlaAspThrCysGlyGlnAspProProLysValProAlaValProA 911  
QY 2113 AGATGAATCATTCATCAATGTCCAGAGAACAG----- 2147  
Db ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :  
911 rgValAsnHisLysAspSerSerIleProLysGlnArgValGluGlyLysGlySerA 931  
QY 2148 -----ATTGCCAGTCTTTTACGTAATGCTTTC-----TTCGCA 2181  
Db ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :  
931 rgSerSerSerAlaAspLysGlySerSerGlyAspThrAlaAsnProPheProValProS 951  
QY 2182 CGTTTCCACGCGCATGCCAAGATCAATCAGATGATTTCCAGTTATCCAGTATTTAACT 2241  
Db ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :  
951 erLeuProAsnGlyAsnSerLys-----ProGlyLysProGlnValLysP 966  
QY 2242 TCAATCGGTGTTGAAGGACGTCATCAAGGAAACACAGAGAGCTTAAACGCTCTTCT 2301  
Db ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :  
966 heAspLysGln-GlnAlaAspLeuHisMetArgGluAlaLysLysMetLysGlnLysAla 985  
QY 2302 GCTACTTTTGAAGAGTCAAGAGAAAAAACCCACTGGGTGGTGACATTCACAGAGACAG 2361  
Db ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :  
:::



Db 805 TrpGlyArgGlnArgGluGlnGlnAspThr-----AlaGluSerSerAspAspPheGlySer 823  
 QY 888 TTTTTCACACATGTAAGCTTGGAAAT----- 914  
 Db 824 GlnLeuArgHisAspTyrValGluAspSerSerGluGlyGlyLeuSerProLeuProPro 843  
 QY 915 -----GCCAAGCAGACGATGGAAGATGAACAG----- 941  
 Db 844 GlnProProAlaArgAlaAspMetThrAspGluGluPheMetArgArgGlnIleLeuGlu 863  
 QY 942 -----GGCAGAGAAGCC 953  
 Db 864 MetSerAlaGluAspAsnLeuGluGluAspThrAlaValSerGlyArgGlyLeu 883  
 QY 954 AGAAGCCAC-----CAGAAGTGTGGC----- 974  
 Db 884 AlaLysHisSerAlaGlnLysAlaSerAlaArgProArgProGluSerSerGlnGluPro 903  
 QY 975 AAGGCTTGCCATCTGCAGAACGCTGTGCAGAGGTGTGCAGAGGAGAGACAGACGTGGTG 1034  
 Db 904 LysArgArgLeuProHisAsnAlaThrThrGlyTyrGluGluLeuLeuSerGluAlaGly 923  
 QY 1035 TCCGAGAGCCCTGTGCGACACTGCTCTGAGGATGTTGGTACTGGACTGAAAAATGCC 1094  
 Db 924 ProAlaGluProThrAspSerSerGlyAlaLeuGlnGlyLeuArgArgPheLysThr 943  
 QY 1095 AACAGATTGAATAGA-----CAAGAAGTAGTCTAGGAATTCCTCTCCA 1139  
 Db 944 IleGluLeuAsnSerThrGlySerTyrGlyHisGluLeuAspLeuGlyGlnGlyProAsp 963  
 QY 1140 TTTCAGAAAGAAAGTGAACCTGAGTCCCAATCGATGATAGATAATCCAAAAATAGTTGT 1199  
 Db 964 ProAsnLeuAspArgGluProGlu-----LeuGluMetGluSerLeuThrGlySerPro 981  
 QY 1200 CAGAGTTCAGAACAGACATGAAGACAAAGT-----CCAGTGTGTTGAT 1241  
 Db 982 GluAspArgSerArgGlyGluHisSerSerThrLeuProAlaSerThrProSerTyrThr 1001  
 QY 1242 GAA-----CAGAAAGATAGACGATTCAGTGTGAGAGCGGTCC 1271  
 Db 1002 SerGlyThrSerProThrSerLeuSerSerLeuGluGluAspSerAspSer----- 1019  
 QY 1272 GCAAAATAACCTTCAGGTTCCACACAGAGAGAGCTGACACTGATGTAGAGAGCGGTCC 1331  
 Db 1020 -----ProSerArgArgGlnArgLeuGluGluAlaLysGlnGlnArgLysAlaArg 1036  
 QY 1332 TCTGCTAAGGGAGTGTGAGATTCGATTACATTTCCAATTTGAAGGAGGAGAGTCCGAGCT 1391  
 Db 1037 HisArgSerHisGlyProLeuLeuProThrIleGlu----- 1048  
 QY 1392 GGAATGAATGATGTGAATGCTCAACAGACCTGGAGTAGTCTTACCTGAGTATGTAGAGTGC 1451  
 Db 1049 -----AspSerSerGluGluGluGluLeu 1056  
 QY 1452 AGAAATCTTAAGCAACATGGGAGAAAGATTCTAAATATCACAGAT---CATTTTCATGAGA 1508  
 Db 1057 ArgGluGluGluGluLeuLeuArgGluGlnGlnLysMetArgGluValGluGlnArg 1076  
 QY 1509 GTG---CCCAAGCAGAGGACAAAGAAAGAAACAATGTGAATGAACATCAAGAACAA 1565  
 Db 1077 IleArgSerThrAlaArgLysThrArgArgAspLysGluGluLeuArgAlaGlnArg--- 1095  
 QY 1566 GAAAGGAGATCCCTAATAATCATTTCCACCTCACTTCTCCAGATAGAAATGCTTGGGA 1625  
 Db 1096 ArgArgGluArgSerLysThrProProSerAsnLeuSerProIleGluAsp-----Ala 1113  
 QY 1626 ACTCCTATTGAGGAGATGAGAGAAATGCAAGGTGTGGATCCGCTGCTCCCTCTGAGA 1685  
 Db 1114 SerProThrGluGluLeuLeuArgGln-----AlaAlaGluMetGluGluLeuHis 1129  
 QY 1686 CCATCTGCCAATCACAGTACTATTTCGGGTAGATCTTTTTCGAATAGGAGAAAGTTCCCT 1745  
 |||||

Db 1130 ArgSerSer----- 1132  
 QY 1746 AAACCTTTCCCAACACACATTTTAAAGATTGTGGGACAAACAGCATGTTAAGATGCTTGT 1805  
 Db 1133 -----Cys 1133  
 QY 1806 TCAGAACAAAACCTTGACCTGTGTGGAAGATGAGAATGGTGTGACGAGCTGCAGCAGCCGG 1865  
 Db 1134 SerGluTyrSerProSerProSerLeuAspSerGluAlaGluThrLeuAspGlyGly--- 1152  
 QY 1866 TGGGAACACTNTTCAGACTGCATTTCTCAACAGGCTCACTCGGCCCCAGAACCTCGAAGGAT 1925  
 Db 1153 -----ProThrArgLeuTyrLys 1158  
 QY 1926 GCTATTTCGAAGTACATAGTGT---GCATATTCTAAGAAATGGGACTTTTACA----- 1973  
 Db 1159 SerGlySerGlyTyrAsnLeuProAlaPheMetSerLeuTyrSerProThrGluThrPro 1178  
 QY 1974 -----GCTTTGATTGATTTCTGGGATAGAGTACTAGAAAGACAGACAGCTCAACAC 2024  
 Db 1179 SerGlySerSerThrThrProSerSerGlyArgProLeuLysSerAlaGluGluAlaTyr 1198  
 QY 2025 TTGTATCAGTCCATCTTGCTGATATGGTGAATTTGCATCTGTCTGCAAAATATTGT 2084  
 Db 1199 -----GluAspMetMetArgLysAla----- 1205  
 QY 2085 ACCAGCCAATACCCTCCTGAAACAGAAAGATGAATCATTCATCATCAATGTCCAGGAA 2144  
 Db 1206 -----GluMetLeuGlnArgGlnGlnGly 1213  
 QY 2145 CAGATTGCCAGTCTTTTA-----GCTAATGCTTTCTTGTGCACGTTTCACACAGC 2195  
 Db 1214 GlnValAlaGlyAlaArgGlyProHisGlyGlyProSerGlnProThrGlyProArgSer 1233  
 QY 2196 AATGCCAAGATGAATCAGACTATTCAGTATTCAGATATTAATCTTCAATCGTGTGTT 2255  
 Db 1234 GlnGlySerPhe-----GluTyrGlnAspThrGlnAspHisAspTyrGly----- 1248  
 QY 2256 GAAGGACGTTTCATCAAGGAAACCAGAGAAAGCTTTAAACCGCTCTCTGCTACTTTAGAAGA 2315  
 Db 1249 ---GlyArgAlaSerGlnPro----- 1254  
 QY 2316 GTCACAGAGAAAACCCACTGGGTG-----GTGACATTACCA 2354  
 Db 1255 ValAlaGluSerThrProAlaGlyLeuGlyAlaAlaValTyrGluGluIleLeuGlnThr 1274  
 QY 2355 AGACAGAGCTTGAA-----GATTTTCCAGAGTGG 2384  
 Db 1275 SerGlnSerIleAlaArgMetArgGlnAlaSerSerArgAspLeuGlyPheThrGluAsp 1294  
 QY 2385 GAAAGATGTGAAAACTCCTGACTCGACTCGACTGTCATGTCACTTACGAAGTACCATAGAAGA 2444  
 Db 1295 LysLysLysGluLysGlnPheLeuAsnAlaGluSerAlaTyrMetAspProMetLysGln 1314  
 QY 2445 AACGGC-----CAGGCGATGTACAGGTGGATTTT 2474  
 Db 1315 AsnGlyGlyProLeuThrProGlyThrSerProThrGlnLeuAlaLaProValSerPhe 1334  
 QY 2475 GCAACCGTTTCGTGGAGTGTGTAAACGAGTGCAGGACTTGTGCAAGAAAGAAATCCCG 2534  
 Db 1335 SerThrSer-----ThrSerSerAspSerSerGlyGlyArgValIleProAspValArg 1352  
 QY 2535 TTTTAAATCAACCCCTGAGTTGATTGTTTTCACGCGCTCTTCACTCAGGTGTGGATCACAAT 2594  
 Db 1353 -----ValThrGlnHisPheAlaLysGluProGlnAsp-ProLe 1365  
 QY 2595 GAATGCTTATATCATCAGGTACTGACAGTACAGTGAATACACAGCTATGCCGAACA 2654  
 Db 1365 uLysLeuHisSerSerProValSerSerThrLeuThrSerLysGluValGlyMetThrPh 1385  
 QY 2655 TACCGTGGCCCGGAGCCATGAAGCAGGAGCGAAGGAGGAGGAGGAGGAGGAGGAGGAG 2714  
 Db 1385 eSerGlnGlySerProAlaThrThrAlaSerProThr-----ArgGlyTyrMe 1403  
 |||||

```

Qy 2715 ACTGAGATCGTCGCATCGAGCCCTCCACTTCAGACGCTACCTCGACCAGTTTG 2769
Db 1403 tThrProThrSerProAlaGlySerGluArgSerProSerThrSerSerThrIle 1421
RESULT 7
TI3564
microtubule-associated protein homolog - fruit fly (Drosophila melanogaster)
N;Alternate names: hypothetical protein EG:49E4.1
C;Species: Drosophila melanogaster
C;Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 17-Nov-2000
C;Accession: TI3564
R;Spanos, L.; Papagiannakis, G.; Siden-Kiamos, I.; Louis, C.
submitted to the EMBL Data Library, April 1999
A;Description: Sequencing the distal X chromosome of Drosophila melanogaster.
A;Reference number: Z17689
A;Accession: TI3564
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-5327 <SPA>
A;Cross-references: EMBL:AL031128; PIDN:CAA20006.1
C;Genetics:
A;Cross-references: FlyBase:FBgn005392
A;Introns: 24/2; 52/3; 104/3; 179/1; 232/1; 1669/3; 2566/1; 4798/3; 5272/1
A;Note: EG:49E4.1
C;Superfamily: Drosophila 576K microtubule-associated protein homolog

Alignment Scores:
Pred. No.: 0.00041 Length: 5327
Score: 175.50 Matches: 229
Percent Similarity: 35.88% Conservative: 180
Best Local Similarity: 20.09% Mismatches: 484
Query Match: 2.42% Indels: 247
DB: 2 Gaps: 45

US-09-302-812-1 (1-4070) x TI3564 (1-5327)
Qy 64 AAGAGGGAAGCAGCGT-----CTGGATAGGCGCTGTTCCGG 102
Db 1448 LysAlaGluSerArgArgGluSerIleAlaLysThrHisLysAspGluSerSerLeuAsp 1467
Qy 103 AGGCTGTGACAGCAGCAGTGCAGAACGAGTCAGCGGCGAGAGGGGCGATGTGCGG---- 158
Db 1468 LysAlaLysGluGlnGluSerArgArgGluSerLeuAlaGlu--SerIleLysProGluS 1487
Qy 159 --GGAGGCCAGCAGAGGGGCGG-----CAGTCCGTC 189
Db 1487 erGlyIleAspGluLysSerAlaLeuAlaSerLysGluAlaSerArgProGluSerValT 1507
Qy 190 CTCGCCAGGTAGTGAATGAGGCTCTAGCCCGGCTGCGCGGAGACTCAGTGTGCGG 249
Db 1507 hrAspLysSerLysGluProSerArgArgGluSerIleAlaGluSerLeuLysAlaGluS 1527
Qy 250 GT-----CCGACATGATGTCGGCGCGCGCTGTGAGCCCTGCACCAAGCGACCCCGCT 303
Db 1527 erThrLysAspGluLysSerAlaProProSerLysGlu-----AlaSerArgPro---- 1543
Qy 304 GGGAGCGGCTGCAACTCTCGCGCGCGCGCTCGAGCGCGCGAGCTTCCCGCGCAGGC 363
Db 1544 --GlySerValValGluSerValLysAspGluThrGluLysSerLysGluProSerArgA 1563
Qy 364 AGAGCGGCTCTCGATCCAGGAGCGCTCCGCTGCGTTCAGTTCAGG----- 407
Db 1563 rgGluSerIleAlaGluSerAlaLysProProIleGluPheArgGluValSerArgProG 1583
Qy 408 -----GTCCCGCGCTCGTTCAGTTCAGTTCAGG----- 432
Db 1583 luserValIleAspGlyIleLysAspGluSerAlaLysProGluSerArgArgAspSerP 1603
Qy 433 CCTCGGCGCGGCGGACAGCACCGAGGCGCGCCACCTCTCTCTTTTTCACACAGAGA 492
Db 1603 roLeuAlaSerLysGluAlaSerArgProGluSerValLeuGluSerValLysAspGluP 1623

493 CTATAACCACTGTGGATGGACACTAAAGAAATCAAGACAGTGTGAATCAGAAAGTTTGCATA 552
1623 rolleLysSerThrGluLysSerArgGluSerValAlaGluSerPheLysAlaAspS 1643
553 GTAAGAAAACAACAT-----ACAAGAGAAATCCATGA 588
1643 erThrLysAspGluLysSerProLeuThrSerLysAspIleSerArgProGluSerAlaV 1663
589 TGAGTCTGTACAAAAGATAACTTTTATCAACATACATCGAA----- 632
1663 alGluAenValMetAspAlaProPheLysGluThrSerArgProGluSerAlaValGlyS 1683
633 -----AAATTAGAAAATGTTTCTCAGCTAGCTAGTGTGATAGTCAACAGTTGAAAAGGT- 686
1683 erMetLysAspGluSerMetSerLysGluProSerArgGluSerValLysAspGlyA 1703
687 --ACACAGTATTGGAACAGCATCAGACTGCGGCTATGTGTAACTGGCAGNATGAAGGC 744
1703 laAlaGlnSerArgGluThrSerArgProAlaSerValAlaGluSerAlaLysAspGlyA 1723
745 CACACTCAGAACGCTTTTGGAAAGTGAACCTCCAGCGGTAACCTCGTACCAGAGCAGT 804
1723 laAspAspLeuLysGluLeu---SerArgProGluSerThrThr----- 1736
805 TCAGTAATGTATATGTCGATCAGTCGTCCTCCCAAGAGATGATCAGACTCACAAATAGTG 864
1737 -----GlnSerLysGluAlaGlySerIleLysAspGluLysSerProLeuAlaSerG 1754
865 AGGAGCTAGAGATAATCAGCAGCTTTTGTGACACATGTAAGTTCGCAATCAAGCAGA 924
1754 luGluAlaSerArgProAlaSerValAlaGluSerValLys-----AspGluAlaGluL 1772
925 CGATGGAAGATGAACAGCGGACAGAGCC-----AGAAGCCACCAGAGTGTGCA 975
1772 ysSerLysGluGluSerArgGluSerValAlaGluLysSerProLeuProSerLysG 1792
976 AGGCTTGCATCTCTGCAGAACCTGTGCGGGTGTCTCAGCAGGAG----- 1019
1792 luAlaSerArgProAlaSerValAlaGluSerIleLysAspGluAlaGluLysSerLysG 1812
1020 -----GAGACAGACGTGTGTCGAGAGCCCTTGTCTCGAC----- 1055
1812 luGluSerArgArgGluSerValAlaGluLysSerProLeuProSerLysGluAlaSerA 1832
1056 -----ACTGGCTCTGAGGATCTGTGACTGCACTGAAATGCAACACAGATTGAATA 1107
1832 rgProAlaSerValAlaGluSerIleLysAspGluAlaGluLysSerLysGluGluSerA 1852
1108 GACAAGAAAGTAGTCTAGGAAATTCCTCT---CCATTGTGAGAAAGAAAGTGAACCTGAGT 1164
1852 rgArgGluSerValAlaGluLysSerProLeuProSerLysGluAlaSerArgProAlas 1872
1165 CACCAATG-----GATGTAGATAATCCAA---AATAGTTGTCCAGGATT 1206
1872 erValAlaGluSerIleLysAspGluAlaGluLysSerLysGluGluSerArgArgGluS 1892
1207 CAGAAGCAGATGAAGACACCAAGTCCAGGTTTGTGATGAACAGAGAGATAGCAGTTCTGCTC 1266
1892 erValAlaGluLysSerProLeuProSerLysGluAlaSerArgProAlaSerValAlaG 1912
1267 AAACA-----GCAATTAACCTTCAAGGTTCCAAACCAAGAGAGCTCACACTG 1314
1912 luSerIleLysAspGluAlaGluLysSerLysGluGluSerArgArgGluSerValAlaG 1932
1315 AG-----TTGAGGAGCGGCTCTCTGCTAAGGAGGAGGTGAGATTCGATTACATTTC 1365
1932 luLysSerProLeuProSerLysGluAlaSerArgProAlaSerValAlaGluSerIleL 1952
1366 AATTGTAGGAGGAGAGAGT-----CGAGCTG 1392
1952 ysAspGluAlaGluLysSerLysGluGluSerArgArgGluSerValAlaGluLysSerP 1972
1393 GAATGAATGATGTGAATGCCAAACGACCTGGAAGTACTTCT---AGCTGAATGTAGAGT 1449

```

```

Db      1972  roLeuProSerLysGluAlaSerArgProAlaSerValAlaGluSerIleLysAspGluA 1992
      : : : : : ||| ||| ||| ||| ||| : : : : : ||| : : : : :
QY      1450  GCGAAATTTCTAGCAACATGGGAGAAAGATTCT-----AAAATCAGATCAT 1500
      : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Db      1992  laGluLysSerLysGluGluSerArgGluSerValAlaGluLys-SerProLeuPro 2011
      : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
QY      1501  TCATGAGAGTGGCCAAAGCAGACGACAAAAGAAAGAAACAATGTGAATGAAACATCAAA 1560
      : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Db      2012  Ser-LysGluAlaSerArgProAlaSerValAlaGluSerIleLysAspGluAlaGlu 2031
      : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
QY      1561  GAACAGAAAGAGATCCCTAAATACATTCACCTCCACCTTCTCCAGATAAGAAATGCC 1620
      : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Db      2031  sSerLysGluGluSer-----ArgGluGluSerValAl 2042
      : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
QY      1621  TTGGAACTCTATTGAGGAGATGAGGAGATGCCAAGGTGTGGATCCGCTCCCTCCCT 1680
      : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Db      2042  aGluLysSerProLeuProSerLysGluAlaSerArgProAlaSerValAlaGluSerIl 2062
      : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
QY      1681  TGAGACCATCTGCCAATCACACAGTCACTATTCCGGGTAGATCTTTGCCGAATAGGAGAAG 1740
      : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Db      2062  eLysAspGluAlaGluLysSerLysGluGluSerArgGluSerAlaAlaGluLysSe 2082
      : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
QY      1741  TTCCTAAACCTTTCCCAACACATTTTAAAGATTTGTGGGCAACAACAGCATGTTAAGATGC 1800
      : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Db      2082  r-ProLeuPro-----SerLysGluAlaSerArgp 2092
      : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
QY      1801  CTTGTTTCAGAACAAACTTGTACCTGTGGAAGATGAGATGTGTGAGCGAGCTGCAGGCA 1860
      : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Db      2092  roAlaSerValAlaGlu-----SerValLysAspGluAlaAspLysSerLysGluGlu 2110
      : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
QY      1861  GCCGGTGGGAACCTCAT-----CAGACTGCAAC 1887
      : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Db      2110  erArgArgGluSerMetAlaGluSerGlyLysAlaGlnSerIleLysGlyAspGlnSerp 2130
      : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
QY      1888  TTCCTAACAGGCTCACTGGCCCCAGAACCTGAAGGATGCTATT----- 1931
      : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Db      2130  roLeuLysGluValSerArgProGluSerValAlaGluSerValLysAspProVal 2150
      : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
QY      1932  -----CTGAAGTACAAATGTCGATTTCTTAAGAAATGGGATTTACAG 1974
      : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Db      2150  ysSerLysGluProSerArgGluSerValAlaGlySerValThraAlaAsp---SerA 2169
      : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
QY      1975  CTTTGTGATGAT-----TTCTGGG 1992
      : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Db      2169  laArgAspGlnSerProLeuGluSerLysGlyAlaSerArgProGluSerValVala 2189
      : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
QY      1993  ATAAGGTACTAGAAAGACGAGAGCTCAACAC-----TTGTATCAGTCCA 2037
      : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Db      2189  spSerValLysAspGluAlaGluLysGlnGluSerArgGluSerLysThrGluSerV 2209
      : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
QY      2038  TCTTGCTGATATGGTGAATTCGACTCTGTCTGCCAAATATTGTACCCAGCCCAATAC 2097
      : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Db      2209  alileProLysAlaLysAspLysSer---ProLysGluValLeuGlnProVal 2228
      : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
QY      2098  CACTC-----CTGAACAGAGATGAATCATTCATCCATCAATGCTCACAG---GAAC 2145
      : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Db      2228  erMetThrGluThrIleArgGluAspAlaAspGlnProMetLysProSerGluAlaGlu 2248
      : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
QY      2146  AGATTGCCAGCTTTTACGTAATGCTTTCTTCGTCAGCTTTCCACGA---CGCAATGCCA 2202
      : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Db      2248  erArgArgGluSerIleAlaGluSerIleLysAlaSerProArgAspGluLysSerP 2268
      : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
QY      2203  AGATGAATCAGACTATTCAGTTATCCA-----GATA 2235
      : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Db      2268  roLeuAlaSerLysGluAlaSerArgProGlySerValAlaGluSerIleLysTyrAspL 2288
      : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
QY      2236  TTAACCTTCAATCGGTGTTTGAAGGAC-----GTTTCATCAAGGAAACCAAGAGAGC 2286
      : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Db      2288  euAspLysProGluIle-IleLysAspLysSerThrGluHisSerArgArgGluSer 2307
      : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
QY      2287  TTAAGACGCTCTTCTGCTACTTTTAGAGAGTCAAGAGAAAACCCATCGGTGGTGA 2346
      : : : : : ||| : : : : : ||| : : : : : ||| : : : : :

```

```

Db      2308  LeuGluAspLysSerAlaValThrSerGluLysSerValSerArgProLeuSerValAla 2327
QY      2347  CATTCACAAGACAGAGTCTTGAAGATTTTCCAGAGTGGGAAAGATCTGAAAAACTCCTGA 2406
      : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Db      2328  SerAspHisGluAlaValAlaValAlaGluAspAspAlaLys----- 2341
      : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
QY      2407  CTCGACTGCTATCTTACGAAGGTACCATAGAAGAAACCGCCAGGGCATCTCAGG 2466
      : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Db      2342  -----SerSerIleSerProLysAspLysSerArgProGlyPhe--- 2354
      : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
QY      2467  TGGATTTTCAACCGTTTCGTTGGAGGTGGTGTAAACCACTGCAGGACGTGTGCAGAAG 2526
      : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Db      2355  ---ValAlaGluThrValSerSerProIleGluGluAlaThrMetGluPheSerLysIle 2373
      : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
QY      2527  AAATCCGCTTTTAAATCAACCTCTGAGTTGATTGTTTTCAGCGCTCTTCACTGAGGTCTGG 2586
      : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Db      2374  GluValValGluLysSerSerLeuAlaLeuSerLeuGlnGlySer-----Gly 2390
      : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
QY      2587  ATCACAATGAATGCTTATCATCAGGTA----- 2616
      : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Db      2391  GlyLysLeuGlnThrAspSerSerProValAspValAlaGluLysAspPheSerHisAla 2410
      : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
QY      2617  -----CTGAGCAGTACAGTGAATACACAGGCTATGCCGAAACATACCGCTGGGCC 2667
      : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Db      2411  ValAlaSerValSerThrValThrProThrLeuThrLysProAlaGluAlaGlnIle 2430
      : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
QY      2668  GGAGCATCAAGACAGGAGCGAAAGGAGCAGCTGGCAGAGGCGCAGCACTGAGATCGTCG 2727
      : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Db      2431  GlyAlaAlaLysThr---ValSerSerProLeuAspGluAlaLeuArgThrProSerAla 2449
      : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
QY      2728  CCAATCAAGCCCTCCACTTCAGACGCTACCTGACAGCAGTTTGTGCCGAGAGATCAGAC 2787
      : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Db      2450  ProGluHisIleSerArgAlaAspSer-ProAlaGlu---CysAlaSerGluGlu----- 2466
      : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
QY      2788  GGGAGCTTAAACAGGCTTACTGTGGATTCTTCGTCCTCGAGTTTCTTCAGAGAACCTGT 2847
      : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Db      2467  -----IleAlaSerGln-AspLysSerProGlnValLeuL 2478
      : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
QY      2848  CTCGAGTGGCTACAGGAAACTGGGCGTGTGGTGCCTTTGGGGGTGATCTAGACTAAAG 2907
      : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Db      2478  ysGluSerSerArgProAlaTTPValAlaGluSerLysAspAlaAlaGlnLeuLys 2498
      : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
QY      2908  CCTAATACAGATCTCGGCGAGCTGTGCTGCTGAGCTGAGCGAGACGTG 2951
      : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Db      2498  erSerValGluAspLeuArgSerProValAlaSerThrGluIle 2512
      : : : : : ||| : : : : : ||| : : : : : ||| : : : : :

```

## RESULT 8

T34036  
 hypothetical protein B0041.7 - *Caenorhabditis elegans*  
 C;Species: *Caenorhabditis elegans*  
 C;Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 29-Oct-1999  
 C;Accession: T34036  
 R;Rulton, R.; Wohldmann, P.  
 submitted to the EMBL Data Library, April 1997  
 A;Description: The sequence of *C. elegans* cosmid B0041.  
 A;Reference number: Z21466  
 A;Accession: T34036  
 A;Status: preliminary; translated from GB/EMBL/DBJ  
 A;Molecule type: DNA  
 A;Residues: 1-1359 <FULL>  
 A;Cross-references: EMBL:AF000196; PIDN:AAC24256.1; GSPDB:GN00019; CESP:B0041.7  
 A;Experimental source: strain Bristol N2; clone B0041  
 C;Genetics:  
 A;Gene: CESP:B0041.7  
 A;Map position: 1  
 A;Introns: 12/2; 59/2; 248/2; 582/1; 675/1; 733/3; 966/2; 1044/2; 1310/3

## Alignment Scores:

Pred. No.: 0.000655 Length: 1359  
 Score: 175.00 Matches: 141  
 Percent Similarity: 34.73% Conservative: 124  
 Best Local Similarity: 18.48% Mismatches: 284  
 Query Match: 2.42% Indels: 214





Db 657 ThrLysArgGileCysLeuThrGlyThrProLeuGlnAsnAsnLeuMetGluTyrHis 676  
 QY 2448 -----GSCCAGGCGCATGCTA-----CAGTGATTTTGCACACCGCT 2483  
 Db 677 CysMetValAsnPheValLysProGlyLeuLeuGlyThrLysThrGluPheAlaAsnArg 696  
 QY 2484 TTCTGTTGGAGGTGGTAAACAGTACAGTGCAGGAGCTTGTGCAAGAA-----GAAATC 2531  
 Db 697 PheVal-----AsnIlelleAsnArgGlyArgThrLysAspAlaSerProLeuGluVal 714  
 QY 2532 CGCTTTTAA 2540  
 Db 715 SerPheMet 717

RESULT 9  
 A47328  
 natural killer cell tumor-recognition protein - human  
 N;Alternate names: cyclophilin-related NK-tumor recognition protein; natural killer-tumor  
 C;Species: Homo sapiens (man)  
 C;Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 24-Sep-1999  
 C;Accession: A47328  
 R;Anderson, S.K.; Gallinger, S.; Roder, J.; Frey, J.; Young, H.A.; Ortaldo, J.R.  
 Proc. Natl. Acad. Sci. U.S.A. 90, 542-546, 1993  
 A;Title: A cyclophilin-related protein involved in the function of natural killer cells.  
 A;Reference number: A47328; MUID: 93133824; PMID: 8421688  
 A;Accession: A47328  
 A;Status: preliminary  
 A;Molecule type: mRNA  
 A;Residues: 1-1403 <AND>  
 A;Cross-references: GB:104288; NID: g181251; PID: AAA35734.1; PID: g181252  
 A;Experimental source: NK killer cells from adult blood  
 A;Note: sequence extracted from NCBI backbone (NCBIN:122798, NCBIP:122800)  
 C;Genetics:  
 A;Gene: GDB:NKTR  
 A;Cross-references: GDB:137171; OMIM:161565  
 A;Map position: 3p23-3p21  
 C;Superfamily: natural killer cell tumor-recognition protein; cyclophilin homology  
 C;Keywords: alternative splicing; lymphocyte  
 F;60-230/domain: cyclophilin homology <CYP>

Alignment Scores:  
 Pred. No.: 0.000715 Length: 1403  
 Score: 174.50 Matches: 158  
 Percent Similarity: 33.84% Conservative: 110  
 Best Local Similarity: 19.95% Mismatches: 298  
 Query Match: 2.41% Indels: 226  
 DB: 1 Gaps: 32

US-09-302-812-1 (1-4070) x A47328 (1-1403)

QY 113 AGCAGGAGCTGCAGAACGAGTCAGCGCAGAGGGGGCATGTGCGCGGAGCGACCGAGGA 172  
 Db 566 ThrHisSerArgAspSerTyrArgSerLysSerHisSerGlnSerTyrSerArgGly 585  
 QY 173 GGGGGCGCAGTCGCTCCCTCCAGGTAGTGAATAGAGGTCTACGCCCGCGGTGGCCCG 232  
 Db 586 -SerSerArgSer-----ArgThrAlaSerLysSerSerHisSerArgSerAr 602  
 QY 233 GAGACTCAGTCTCGGGTCCACATGAGTCGCGGCGCCCGCTGTGAGCCCTGCACCAA 292  
 Db 602 gSerLysSerArg-----SerSerLysSerGlyHisArgLysArgAlaSerLy 619  
 QY 293 GCGACCCCGCTGGGACGCGCGCT-----GCAACTCTCCGCC 328  
 Db 619 sSerProArgLysThrAlaSerGlnLeuSerGluAsnLysProValLysThrGluProLe 639  
 QY 329 GGCGCGCTCGGACCGCGGAGC-----TTCGCCGCGAGCGAG 367  
 Db 639 uArgAlaThrMetAlaGlnAsnGluAsnValValGlnProValAlaGluAsnIl 659  
 QY 368 GCGGCTCTCGATTCAGGAGCGCTCCGGTCAGTTCCAGGTCCCGCGCTCGTCAGG 427  
 Db 659 eProValIleProLeuSerAspSerProProSerArgTrpLysProGlyLysPr 679

QY 428 CTGC-----GCCCTGGCGCGCGGACAGCAGCCAGGCGCGCCACCTCTCTCTTTT 481  
 Db 679 oTrpLysProSerTyrGluArgIleGlnGluMetLysAlaLysThrHisLeuLeuPr 699  
 QY 482 CAAACAGAGACATATAACCCAGTTGGATGCACACTAAGCAATCAACAGCAGTTGAATCAGA 541  
 Db 699 oIleGlnSerThrTyr---SerLeuAlaAsnIleLys-----GluThrGlySerSerSe 716  
 QY 542 AAGTTTGCACTAGTAAACAAACAAATACAGAGAGAAATCCATGATGATGATTCTCTGA-- 599  
 Db 716 rSerTyrHisLysArgGluLysAsnSerGluSerAspGlnSerThrTyrSerLysTyrSe 736  
 QY 600 -----CAAAAGATTAACITTTA 616  
 Db 736 rAspArgSerSerGluSerSerProArgSerArgSerArgSerArgSerArgSerTyr 756  
 QY 617 TCAACATACATAGGAAAAATTAGAAAAATGTT---TCTCAGCTAGGTTTTCATAGTCACC 673  
 Db 756 rSerArgSerTyrThrArgSerArgSerLeuAlaSerSerHisSerArgSerArgSerPr 776  
 QY 674 AGTTGAAAAGGT-----ACACAGTATTGAAAGCAGCATCAGACTCGGCTATGTG 724  
 Db 776 oSerSerArgSerHisSerArgAsnLysTyrSerAspHisSerGlnCysSerArgSerSe 796  
 QY 725 TAAGTGG-----CAGAAATGAGGGCCACACTCAGAACGGCTTTTGGAAAGTGA 772  
 Db 796 rSerTyrThrSerIleSerSerAspAspGlyArgAlaLysArgLysArgLysSerSe 816  
 QY 773 ACTTCAGCGGTAACTCTGTTACAGCAGCAGTTTCAGTAAATGCTAAATGTC----- 821  
 Db 816 rGlyLysLysAsnSerValSerHisLysLysHisSerSerSerSerSerGluLysThrLeuHi 836  
 QY 822 -----GATCAGTCGTCTCCCAAGAGGATGATCATCAGTCGACACAAA 859  
 Db 836 sSerLysTyrValLysGlyArgAspArgSerCysValArgLysTyrSerGluSerAr 856  
 QY 860 TAGT-----GAGCAGAGTAGATAATCAGCAGTGTTCACACATGTAAGCTTGGCAA 913  
 Db 856 gSerSerLeuAspTyrSerSerAspSerGluGln-----SerSerValGlnAlaThrGl 874  
 QY 914 TGCAAGCAGCAGCATGGNAGATCAACAGGCGCAGAGCCAGAGCCACACAGAGTGTGG 973  
 Db 874 nSerAlaGlnGluLysGlu---LysGlnGlyGlnMetGluArgThrHisAsnLysGlnGl 893  
 QY 974 CAAGGCTTCCCTCCTGCAGAGCCTGTGCAGGGGTGTGCAGCAGGAGAGCAGACGTGGT 1033  
 Db 893 uLys-----AsnArgGlyGluLysSerLysSerGl 904  
 QY 1034 GTCGAGAGCCCTTGTGCGAC-----ACTGGCTCTGAGGATGTTGGTACTGGACTGAA 1087  
 Db 904 uArgGluCysProHisSerLysLysArgThrLeuLysGluAsnLeuSerAspHisLeuAr 924  
 QY 1088 AAATGCCAACAGATTGAATAGACAGAAAGTAGTCTAGGAAATCTCCTCCATTTGAGAA 1147  
 Db 924 gAsnGlySerLysProLysArgLys-----AsnTyrAlaGlySerLysTr 939  
 QY 1148 AGAAATGAACCTGAGTCACCAATGGATGTA---GATAATTCAAAAATAGTTGTCAGGA 1204  
 Db 939 pAspSerGluSerAsnSerGluArgAspValThrLysAsnSerLysAsnAspSerHisPr 959  
 QY 1205 TTCAGAGCAGATCAAGAGACAAGTCCAGGTTTGTAGTGAACAGAGATAGCAGTTCTGC 1264  
 Db 959 oSerSerAspLysGluGluGlyGluAlaThrSerAspSerGluValSerGluIl 979  
 QY 1265 TCAACAGCAAAATAACCTTCAAGGTTCCAAAGAGAAAGCTGACACTGAGTTGAGAA 1324  
 Db 979 eHisIleLysValLysProThr-----Th 987  
 QY 1325 GCGGTCTCTGCTAAGGAGGTGAGATTCGATTACATTTCOATTTGAAGAGGAGAGAG 1384  
 Db 987 rLysSerSer-----Th 991



Db 397 erValThrGlnAsnGlnLysGln-----TyAspThrSer-----SerLysThrH 412  
 QY 691 AGTATTTGAAGCAGCATCAGACTGCGCTATGTGAAG-----TGCCAGA 735  
 Db 412 isSerAsnSerGlnGlnGlyThrSerSerMetLeuGluAspLeuGlnLeuSerAsps 432  
 QY 736 ATGAGGGCCACACTCAGAACCGCTTTTGAA-----AGTGAACCTCCA- 779  
 Db 432 erGluAspSerAspSerGluGlnThrProGluLysProSerSerSerAlaProProS 452  
 QY 780 --GCGTAACCTCGTACAGACAGCTTCAGTAATGCTAAT----- 818  
 Db 452 erAlaProGlnSerLeuProGluProValAlaSerAlaHisSerSerSerAlaGluSerG 472  
 QY 819 -----GTCCGATAGTCGTCCCCAAGAGATGATCAGAGTGACACAAATAGTGAGAGA 870  
 Db 472 luSerThrSerAspSerSerSerSerSerSerSerSerSerSerSerSerSerSerAsps 492  
 QY 871 GTAGAGTAATCAGCAGTCTTTTGCACATGTAAGCTTCGGAATGCCAAGCAGCAGTGG 930  
 Db 492 erGluGlnAsnGlu-----ProLeuGluThrProAlaProGluProG 506  
 QY 931 AAGATGAACAGGCGACAGAGCAGAGCCAGAGGCTGTGCGAAGGCTTGCATCCTCG 990  
 Db 506 luProProThrThrAsnLysTrpGlnLeuAspAsnTrpLeuThrLysValSerGlnProA 526  
 QY 991 CAGAAAGCTGTGAGGGTGTGACAGAGGAGCAGAGCGTGTCGCGAGAGCCCTTGT 1050  
 Db 526 laAlaProProGluGly-----ProArgSerThrGluProProArgArgHisProGluS 544  
 QY 1051 CGCACACTGCTGAGGATGTTGCTGACTGGACTGAAATGCCACACAGATTCAATAGAC 1110  
 Db 544 er---LysGlySerSerAspSerAlaThrSer-----G 554  
 QY 1111 AAGAAGTAGTCTAGCAATCTCTCCATTTGAGAAAGAAAGTGAA---CCTGAGTCAC 1167  
 Db 554 lngLHisSerGluSerLysAspProProLysSerSerSerSerLysAlaProArgAlap 574  
 QY 1168 CAATGGATGTAGATAAT---TCCAAATAATGTGTGAGGATTCAGAAAGCAGATGAAGAGA 1224  
 Db 574 roProGluAlaProHisProGlyLysArgSerCysGlnLysSerProAla----- 590  
 QY 1225 CAAGTCCAGGTTTGATGACAGGAGATAGCAGTCTGCTCAACAGCAGCAATAACCTT 1284  
 Db 591 -----GlnGlnGluProProGlnArgGlnThrValGlyThrLysG 604  
 QY 1285 CAAGGTTCCAAACAGAGAGCTGACACTGAGTTGAGGAAGCGTCTCTGCTAAGGGAG 1344  
 Db 604 lnProLysLysProValLysAlaSerAlaArgAlaGlySerArgThrSerLeuGlnGly- 623  
 QY 1345 GTGAGATTGATTACATTTCCAAATTTGAGGAGGAGAGAGTTCGAGCTGGAATGAATG 1404  
 Db 624 -----GluArgGluProGlyLeuLeuProTyGlySerArgAspG 637  
 QY 1405 TGAATGCCAAA---CGACCTGGAAGTACTTCTAGCCCTGAATGATGAGTGAGCAAAATCTA 1461  
 Db 637 lnThrSerLysAspLysPro-----LysValL 646  
 QY 1462 AGCAACATGGGAGA---AAGGATTCTAAAATCACAGATCATTTTCATGAGAGTGCCCAAG 1518  
 Db 646 ysThrLysGlyA-gProArgAlaAlaSerAsnGluProLysProAlaValProProS 666  
 QY 1519 CAGAGCAGAAAAGAAAGAAACAATGT----- 1544  
 Db 666 erSerGluLysLysLysHisLysSerSerLeuProAlaProSerLysAlaLeuSerGlyP 686  
 QY 1545 --GAATGAACATCAAGAACAGAGAGGAGATCCCTAATACAT----- 1589  
 Db 686 roGluProAlaLysAspAsnValGAspArgThrProGluHisPheAlaLeuValProL 706  
 QY 1590 -----CCACCTCACCTTCTCCAGATAAGAAATGGCTTGA----- 1625

Db 706 euThrGluSerGlnGlyProProHisSerGlySerGlySerArgThrSerGlyCysArgG 726  
 QY 1626 -----ACTCCTATTGAGGAGATGAGCAATGCCAAGGTGTGGGATC----- 1667  
 Db 726 lnAlaValValGlnGluAspSerArgLysAspArgLeuProLeuArgAspT 746  
 QY 1668 -----CGGCTGCTCCCTTGACA-----CCATCTGCCAATCACACAGTGACTATTTC 1713  
 Db 746 hrLysLeuLeuSerProLeuArgAspThrProProGlnSerLeuMetValLysIleT 766  
 QY 1714 GGTAGATCTTTTTCGAATAGGAGAAGTTCCTAAACCTTTCCCAACACATTTTAAAGATT 1773  
 Db 766 hrLeuAspLeuLeu-----SerArgIleProGlnPro----- 776  
 QY 1774 TGTGGGACAAACAGCATGTTAGATGCTCTGTTTCAGAACAAACCTTGATCCCTGTG--G 1830  
 Db 777 -----ProGlyLysGlySerArgGlnArgLysAlaGluAspLysGlnProProAlaGlyL 795  
 QY 1831 AAGATGAGAATGTCGAGCGAGCTGCAGGCGAGCGCGGTGGGAACCTCATTAGATGCACTTC 1890  
 Db 795 ysLysHisSerSerGluLysArgSerSerSerSerSerSerLysLeuAlaLysLysArgL 815  
 QY 1891 TCACAGGCTCACTCGGCCCCAGAACCTGAGGATGCTATTCTGAAGTACATATGGCAT 1950  
 Db 815 ysGlyGluAlaGluArgAspCysAspAsnLysLysIleArgLeuGluLysGluLysLys 835  
 QY 1951 ATTCTAAGAAATGGGACITTCAGACTTGTGATTGATTCTGCGATTAAGGTACTAGAGAAG 2010  
 Db 835 erGlnSerSerSerSerSerHisLysGluSerSerLysThrLysProSerArgP 855  
 QY 2011 CAGAAAGCTCAACACTTGTATCAGTCCATCTTCCTGATATGTTGTAATAATGCACTGTCT 2070  
 Db 855 roSerSerGlnSerSerLysLysGluMetLeuPro----- 867  
 QY 2071 TCCCAAAATATTGTACC-----CAGCCAATA---CCACTCTGAAACAGAGATGA 2118  
 Db 868 ---ProProValSerSerSerGlnLysProAlaLysProAlaLysArgSerArgA 887  
 QY 2119 ATCATTCCATCAATGTCACAGGAA-----CAGATTGCCACTCTTTAGCTAATG 2169  
 Db 887 rgLAlaAspThrCysGlyGlnAspProProLysSerAlaSerSerThrLysSerAsnH 907  
 QY 2170 CTTTCTCTGACAGTTCACGAA-----CGCATGCCAAGATCAAAAT 2211  
 Db 907 isLysAspSerSerLysProLysGlnArgValGluGlyLysGlySerArgSerS 927  
 QY 2212 CAGAGTATTTC----- 2222  
 Db 927 erGluHisLysGlySerSerGlyAspThrAlaAsnProPheProValProSerLeuProA 947  
 QY 2223 -----AGTTATCCAGATATTAACTTCAATCGTGTGTTGAAGGACGTT 2265  
 Db 947 snGlyAsnSerLysProGlyLysProGlnValLysPheAspLysGln-GlnAlaAspLeu 966  
 QY 2266 CATCAAGGAAACACAGAGACTTAAACGCTCTCTGCTACTTTAGAGAGTGCACAGAGA 2325  
 Db 967 HisMetArgGluAlaLysLysMetLysGlnLysAlaGluLeuMet----- 981  
 QY 2326 AAAAACCCTGGTGTGACATTCAAGAGCAGAGCTTTCAGAGATTTTCCAGAGTGGG 2385  
 Db 982 -----ThrAspArgValGlyLysAlaPheLysTyrLeu 992  
 QY 2386 AAAGATGTCAAAACTCCTGACTCGACTCATGCTACTTACGAAGGTACCATAGAGGAA 2445  
 Db 993 GluAlaValLeu---SerPheIleGluCysGlyIleAlaThrGluSerGlu---SerGln 1010  
 QY 2446 AGCGCCAGGCGATCTACAGTGGGATTTTGCAAACCGTTTCGTTG----- 2490  
 Db 1011 SerSerLysSerAlaTyrSerValTyrSerGluThrValAspLeuIleLysPheIleMet 1030  
 QY 2491 -----GAGTGGTGTAAACCGTCCAGC----- 2514  
 Db 1031 SerLeuLysSerPheSerAspAlaThrAlaProThrGlnGluLysIlePheAlaValLeu 1050

```

Qy 2515 -----TTGTGCAAGAGAAATCCGCT 2535
Db 1051 CysMetArgCysGlnSerIleLeuAsnMetAlaMetPheArgCysLysLysAspIleAla 1070
Qy 2536 -----TTTTAATCAACCTCAGTGTGATTGTTTTCACGGCTCTTTCACCTGAGGTGCTGGATC 2589
Db 1071 IleLysTyrSerArgThrLeuAsnLysHisPheGluSerSerLysValAlaGlnAla 1090
Qy 2590 ACAATGAATGCTTATCATCATCAGGTACTGAGCAGTACAGTGAATACACAGGCTATGCGG 2649
Db 1091 ProSerProCysIleAlaSerThrGlyThrProSerProLeuSerProMetProSerPro 1110
Qy 2650 AAACATACCGCTGCGCGGAGCCATGAGACAGAGCGAAGGACGACTGGCAGAGGC 2709
Db 1111 AlaSerValGly---SerGlnSerSerAlaGlySerValGlySerSerGlyValAla 1129
Qy 2710 GCACGACTGAGATCGTCCCATCGACGCGCTCCACT-----TCAGAC 2751
Db 1130 AlaThrIle-----SerThrProValThrIleGlnAsnMetThrSerSerTyr 1145
Qy 2752 GCTACCTCGACAGTTTGTGCGCGAGAGATCAGAC-----GGGAGCTTA 2796
Db 1146 ValThrIleThrSerHisValLeuThrAlaPheAspLeuTyrGluGlnAlaGluAlaLeu 1165
Qy 2797 ACAAGCTTACTGTGGATTCTTC 2820
Db 1166 ThrArgLysAsnLysGluPhePhe 1173

RESULT 11
A46194
neurofilament protein NF-220, high-molecular-weight splice form - longfin squid
C:Species: Loligo pealeii (longfin squid)
C>Date: 22-Sep-1993 #sequence_revision 25-Apr-1997 #text_change 17-Mar-2000
C:Accession: A46194
R:Way, J.; Hellmich, M.R.; Jaffe, H.; Szaro, B.; Pant, H.C.; Gainer, H.; Battey, J.
Proc. Natl. Acad. Sci. U.S.A. 89, 6963-6967, 1992
A:Title: A high-molecular-weight squid neurofilament protein contains a lamin-like rod
A:Reference number: A46194; PMID:92357751; PMID:1379729
A:Accession: A46194
A>Status: nucleic acid sequence not shown; not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-1200 <GW>
A:Cross-references: GB:M94389; NID:g161291; PID:g161292
A:Experimental source: stellate ganglion
A>Note: sequence extracted from NCBI backbone (NCBIP:113499)
C:Superfamily: neurofilament triplet H protein
C:Keywords: alternative splicing

Alignment Scores:
Pred. No.: 0.0013 Length: 1200
Score: 170.50 Matches: 116
Percent Similarity: 36.35% Conservative: 85
Best Local Similarity: 20.98% Mismatches: 257
Query Match: 2.35% Indels: 95
DB: 2 Gaps: 21

US-09-302-812-1 (1-4070) x A46194 (1-1200)
Qy 178 CGCAGTCCGTCCTCCACGAGTTAGTGAATGAGGCTCTACGCCCGGCTGGCCCGGAGAC 237
Db 644 LysSerProLysSerProGlySerLysArgAspThrLeu-----PheGlnSerAsp 660
Qy 238 TCAGTCTCGGGTCCACGATGAGTGGCGGCGCCCGCTGTG----- 279
Db 661 LysValThrSerProValLeuLeuGluArgAlaMetSerLeuGlnCysAlaProThrGlu 680
Qy 280 -----ACCCCTGCACCAAGCGACCGCCGCTGGGACGCGCTG 315
Db 681 LysProLeuSerAspThrIleLysSerProValValSerGluProAlaIleSerProVal 700
Qy 316 CAACCTTCTCCGCGCGCGCTCGGACGCCCGCGAGCT-----TCCCGCGGAGGC 363

```

```

Db 701 SerSerValArgSerGlyAlaGluThrSerProAlaGluArgMetThrSerProIleGly 720
Qy 364 AGAGGCGCGTCCCTCGATTCCCAAGAGCGCTCCGGTGCAGATTCCAGGTCGCCGCGCTCCGCT 423
Db 721 SerGlySerGluLysSerAlaLysSerProValArgSerGluAlaThrLysSerPro---- 739
Qy 424 CAGCTCGCGCTCGGCGCGCGGAGCAGCAGCCAGGAGCGCCACCTCTCTCTTTTCA 483
Db 740 ValSerGluLysSerAlaGlySerLysSerProValPro-----Ser 755
Qy 484 AACAGAGACTATACCACTGGTGGATGACACTAAAGGAATCAACACAGTTGAATCAGAAA 543
Db 756 GluArgAlaLeuSerProVal-----LeuSerGlu--MetSerValHisSerThr 772
Qy 544 GTTTG---CATAGTAAAGAAACAAACAATACACAGAGAAATCATCATGATGATGTTCTGTAC 600
Db 772 laMetSerHisThrSerArgSerProThrAlaSerGluLysSerValLysSerProThrH 792
Qy 601 AAAAAGATAACTTTTATCAACATAAACATGGAATA-----TTAGAAATGTTTCTC 651
Db 792 isSerGluArgThrAlaSerProThrAlaLysSerProIleIleMetGluProAla---- 810
Qy 652 AGCTAGTTTGTGATAAGTCACTGACCTTGAAGAGGTACACAGTATTGAAAGCAGCATCAGA 711
Db 811 -----LysSerProLysAspGluSerGluLysGluSerProGluLys 826
Qy 712 CTGCGCTATGTGTAAGTGGCAGAAATGAA-----GGGCCACACT 750
Db 826 erGluSerValMetSerGlyLysSerGlnIleThrSerSerSerAlaLysSerProValP 846
Qy 751 CAGAACGGCTTTTGGAAAGTGAACCTCCAGCGGTAACTCTGTACCAAGCAGAGTTCAGTA 810
Db 846 roSerGluLysAlaAspSerGluLysSerAlaThrSerProThrProSerGlu-----L 864
Qy 811 ATGCTAATGTGCATCAGTCGTCGCCCAAGAGTATGATCAGTGAC---ACAAATAGTAGG 867
Db 864 ysValAspGluSerSerAlaArgSerProThrHisSerAspArgSerGluSerGlnL 884
Qy 868 AGAGTAGAGATTAATCAGCAGTGTGTTTGCACATGTAAAGCTTGCAGATCAAGCAGACGA 927
Db 884 ysserAlaArgSerProMetThrSerAspHisIleLysSerProIleAspPheGluLys 904
Qy 928 TGAAGATGAACAGGGC----- 944
Db 904 laGluSerGluLysSerAlaLeuSerTyrLysSerSerAspHisAspLysSerProValP 924
Qy 945 --AGAGAGCCAGAGCCAGCCAGAGTGTGCGAAGCTTGCATCTCCAGAGAGCTGTG 1002
Db 924 roSerGluLysAlaGluSerGluLysSerAlaArgSerProValProSerGluLysAlaG 944
Qy 1003 CAGGCTGTACAGCAGGAGGAGACAGACGTCGTGTCGAGAGCCCTTGTCCGACACTGGCT 1062
Db 944 luSerGluLysSerAlaArgSerProValThrSerGluProAlaLysSerProValProS 964
Qy 1063 CTGAGGATGTTGGTACTGCACTGAAAATGTC-----AACGATTGAATA 1107
Db 964 erGluLysAlaAlaGluSerGluLysSerAlaArgSerProValProSerGluLysAlaG 984
Qy 1108 GACAAGAAAGTAGTCTAGGAATCTCTCCATTGTGAGAAA---GAAAGTGAACCTGAGT 1164
Db 984 luSerGluLysSerAlaArgSerProValProSerGluLysAlaGluSerGluLysSerA 1004
Qy 1165 CA-----CCATGATGTAGATAATTCAAAATAATAGTTGTGTCAGATTGAGAG 1212
Db 1004 laArgSerProValProSerGluLysAspGluSerGluLysSerAlaLysSerProThrS 1024
Qy 1213 CAGATGAAGACAGATCCAGTTTGTATGACAGGAGATAGCAGTTCGTCTCAACAG 1272
Db 1024 erAspGluAlaArgSerPro-----ValProSerGluLysAlaGluSerGluLysSerA 1042
Qy 1273 CAAATAAACCT-----TCAAGGTTCACCAACCAAGAGAGCTGACA 1311
Db 1042 laArgSerProValThrSerGluHisValLysSerProValProSerGluLysSerGluS 1062

```

QY 1312 CTGAGTGTGAGGAGCGCTCTCTGCTAAGGAGGTGAGATTGATTCATTTCCAAATTG 1371  
 Db 1062 erGlulysSerAlaArgSerProValThrSerGluHisVallySerProValProSerG 1082  
 QY 1372 AAGGAGGAGAGAGTGCAGTGGAAATGATGATGTAATGCCAAACGACCTGGAAGTACTT 1431  
 Db 1082 luYsAlaGluSer-----GluYsSerAlaLysSerProValAlaSerG 1097  
 QY 1432 CTGAGTGTGAGTGCAGGAATTTCTAAGCAATCGGAGGAGGAGGATTCATAA---- 1487  
 Db 1097 luProAlaLysSerProileProSerGluYsAlaGluSerGluYsSerAlaLysSerP 1117  
 QY 1488 --ATACACATCATTTTCATGAGAGTGCCTAAGCAAGCAGGACCAAAAGA-----AAAG 1536  
 Db 1117 roValThrSerGluHisAlaThrSerProValAlaSerGluYsSerGluYsLys 1137  
 QY 1537 AACAAATGTAATGAACATCAAGAACAGAAAGAGAGATCCCTAAATACATTCACCTC 1596  
 Db 1137 exLysSerProAlaAlaSerGluYsAlaGluSerGluYsSerLysSerProValProS 1157  
 QY 1597 ACCTTCTCCAGTAAGAAATGCTTGGAACTCCT 1631  
 Db 1157 exGluYsGluGlySerLysThrAlaGlySerPro 1168  
 RESULT 12  
 B47328  
 natural killer cell tumor-recognition protein - mouse  
 N:Alternate names: cyclophilin-related NK-tumor recognition protein; natural killer-tumor  
 C:Species: Mus musculus (house mouse)  
 C>Date: 17-Jul-1994 #sequence\_revision 17-Jul-1994 #text\_change 05-Nov-1999  
 C:Accession: B47328; I77662  
 R:Anderson, S.K.; Gallinger, S.; Roder, J.; Frey, J.; Young, H.A.; Ortaldo, J.R.  
 Proc. Natl. Acad. Sci. U.S.A. 90, 542-546, 1993  
 A>Title: A cyclophilin-related protein involved in the function of natural killer cells.  
 A:Reference number: A47328; MUID:93133824; PMID:8421688  
 A:Accession: B47328  
 A>Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-1507 <AND>  
 A:Cross-references: GB:I04289; NID:gl92866  
 A>Note: authors translated the codon AGI for residue 972 as Arg  
 R:Rinfret, A.; Anderson, S.K.  
 Mol. Immunol. 30, 1307-1313, 1993  
 A>Title: IL-2 regulates the expression of the NK-TR gene via an alternate RNA splicing m  
 A:Reference number: 157820; MUID:94019422; PMID:8413330  
 A:Accession: I77662  
 A>Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 235-237;263-294 <RIN>  
 A:Cross-references: GB:S65998; NID:9425701; PIDN:AAB28500.1; PID:9425702  
 C:Genetics:  
 A:Gene: NK-TR  
 C:Superfamily: natural killer cell tumor-recognition protein; cyclophilin homology  
 C:Keywords: alternative splicing; lymphocyte  
 F:60-230/Domain: cyclophilin homology <CYP>  
 Alignment Scores:  
 Pred. No.: 0.00137 Length: 1507  
 Score: 170.50 Matches: 146  
 Percent Similarity: 33.38% Conservative: 91  
 Best Local Similarity: 20.56% Mismatches: 301  
 Query Match: 2.35% Indels: 172  
 DB: 2 Gaps: 31  
 US-09-302-812-1 (1-4070) x B47328 (1-1507)  
 QY 101 GGAGGCTGTGAGACAGAGAGCTGCAGAGAGTGCAGGCGGAGAGGGGGGATGGTCCGGG 160  
 Db 562 GlySerAlaSerThrHisSerArgAspSerLysSerLysSerHisSerArgSer 581  
 QY 161 AGGCACGAGGAGGGGGC-----GCAGTCCGTCCCTCCCGAGGTTAGTGAATG--- 208  
 :|||||  
 :|||||

Db 582 AspSerArgGlySerSerArgSerArgAlaValSerLysSerSerArgSerSerLeuAsn 601  
 QY 209 -----AGCTCTAGCCCGGCTGGCCCGGAGACTCAGTGTGCGGGTCCCGAG 256  
 Db 602 ArgSerLysSerArgSerSerArgSerGlyPro-ArgArgThrSerIleSerProly 621  
 QY 257 CATGAGTGGGGCCCCGGC-----TGTGAGCCCTGCACCAAGCGACC 298  
 Db 621 sLysProAlaGlnLeuSerGluAsnLysProVallyThrGluProLeuArgProSerVa 641  
 QY 299 CGCTGGGAGCGCGCTGCACCTCTCCCGCCCGCCCTCGGACGCCCGAGTTCCTCCCGG 358  
 Db 641 lProGlnHisGlyAsnValLeuValGlnProValAla-----AlaGluAsnIlePro-- 658  
 QY 359 CAGGCGAGGCGCGCTCCGATCCAGAGACGCTCCGGTGCAGTTCAGGTCCTCCCGCTC 418  
 Db 659 -----ValIleProLeuSerAspSerProProSerArgTrpLysProG 674  
 QY 419 CTCGTCCAGGCTGC-----GCCCTGGCGCGGCGGACAGCACCGAGCGCGCCACCTC 472  
 Db 674 yGlnLysProTrpLysProSerTy-GluArgIleGlnGluMetLysAlaLysThrThri 694  
 QY 473 TCTTCTGTTTCAAAACAGAAAGACTATAACCCAGTTGGATGGACACTAAAGGAATCAAG---AC 529  
 Db 694 sLeuLeuProValGlnSerThrTy-Ser-----LeuThrAsnIleLysAlaTh 710  
 QY 530 AGTTGAATCAGAAAGTTTGCATAGTAAGAAACAAACAATAACAGAGAGAATCCATGAT 589  
 Db 710 rValSerSerSerTyHisLysArgGluLysProSerGluSerAspGlySerAlaTy 730  
 QY 590 GAGTCTCTGACAAAAGATACTTTTATCAACATAACATGAGAAATAGAAAATGTTTC 649  
 Db 730 rSerLysTySer----- 734  
 QY 650 TCAGCTAGGTTTGTATAGTCCACAGTTGAAAAGGTACACAGTATTGGAAG----- 701  
 Db 735 -----AspArgSerSerGlySerGlyArgSerGlySerLysSerSerAr 750  
 QY 702 ----CAGCATCAGACTCGCGCTATGTGTAAGTGGCAGAAATGAAGGCCACCTCAGAACG 757  
 Db 750 gSerArgSerSerSerArgSerTyThrArgSerArgSerArgSerLeuProThrSerAr 770  
 QY 758 GCTTTTGGAAAGTGAACCTCCAGCGTAACTCTGTTACAGAGCAGCTTCAGTAATCTAA 817  
 Db 770 gSerLeuSerArgSerProSerSerArgSerHisSerProAsnLysTy-SerAspGlySe 790  
 QY 818 TGTCTGATCAGTCTGCTCCCAAGAGATGATCAGATGACACAATATAGTGGAGAGTAGA-- 875  
 Db 790 r-----GlnHisSerArgSerSerTyThrSerValSerSerAspGlyArgAr 808  
 QY 876 -----GATAATCAGCAGTTTGTGACACATGTAAG-----CTTCGGA 913  
 Db 808 galaMetPheArgSerAsnArgLysLysSerValThrSerHisLysArgHisArgSerAs 828  
 QY 914 TGCAAAAGCAGACATGGAAATGAA-----CAGGCGAGAGAACCCAGAACG---CACCA 964  
 Db 828 nSerGluLysThrLeuHisSerLysTyValArgGlyArgGluLysSerSerArgHisAr 848  
 QY 965 GAAGTGTGGCAGGCTTGCATCTCCTCAGAGCCTGTGCGGCTGTGAGGCTGTGAGGAGGAC 1024  
 Db 848 glyTySerSerGluSerArgSerSerLeuAspTyThrSerAspSerAspGlnSerHisVa 868  
 QY 1025 AGAGTGTGTCTCCGAGAGCCCTTGTGCGGACACTGCTGAGAGATGTTGGTACTGGACT 1084  
 Db 868 lGlnValTySerAlaProGluLysGluLysGlnGlyLysValGluAlaLeuAsnAspLy 888  
 QY 1085 GAAAATGCCAACAGATTGAATAGACAAAGAAAGTAGTCTAGGAAATTCCTCCATTTGA 1144  
 Db 888 sGlnGlyLysGlyArgGluGluGlyLysProLysProGluTrpGluCysProArgSerLy 908  
 QY 1145 GAAAGAAAGTGAACCTGAGTCCACCAATGGATGTAGATTAATCCAAA---AATAGTTGTCA 1201  
 Db 908 sLysGluAsnSerGluAspHisSerArgAspSerValSerLysGlyLysAsnCysAl 928  
 :|||||  
 :|||||

A;title: pCal, a highly unusual Ty1/copia retrotransposon from the pathogenic yeast *Candida albicans*  
A;reference number: Z14877, MUID:98037512; PMID:9371461  
A;accession: T03277  
A;status: preliminary; translated from GB/EMBL/DBJ  
A;molecule type: DNA  
A;residues: 1-1576 (MAT)  
A;cross-references: EMBL:AF007776; NID:g2636718; PIDN:AAC49878.1; PID:g2636719  
C;genetics:  
A;mobile element: retrotransposon pCal

Alignment Scores:  
Pred. No.: 0.00139 Length: 1576  
Score: 170.50 Matches: 157  
Percent Similarity: 33.09% Conservative: 113  
Best Local Similarity: 19.24% Mismatches: 291  
Query Match: 2.35% Indels: 255  
DB: 2 Gaps: 37

US-09-302-812-1 (1-4070) x T03277 (1-1576)  
QY 281 GCCTGCACCAAGCGACCCGCTGGGACGCGCTGCAACTTCTCCGCGCGCGCTCGGA 340  
DB 654 SerLeuHisGluLeuThrPro-GlyAspAsnProValSerLysProGlnLeuGlyTh 673  
QY 341 CGCCCGGAGCTTCCCGCGCAGCAGCGCGCTCTC-----GATTCCAGACGCTCC 394  
DB 673 rGluThrSerValIleGlySerLysGluProIleThrAsnHisThrLysAspAlaPr 693  
QY 395 G---GTCAGTTCAGGTCCTCGCTCGTTCAGGCTGGCGCTGGCGCGCGCGGACGA 451  
DB 693 oSerIleGlnGlyArgAspHisLysArgSerGluSerThrAla-----GlnValGlyLe 711  
QY 452 GCACCGAGGCGCGCCACTCTCTTGTGTTTCAACACAGACACTATAACCCAGTTGGATGA 511  
DB 711 uSerHisGlnProGlnThrGlyThrProAlaSerGluGluSerLysLeuSerGlyThrAs 731  
QY 512 CACTAAAGGATCAAGACAGTTGAA-----TCAGAAAGTTTGATAGTAAGAAAA 562  
DB 731 pHisPheGlyValAspValValLysGluThrValSerGluAspTrpHisThrSerAspTy 751  
QY 563 CAACAATCAAGA-----GAAGATCCATGATGATGCTCTGTACAAAAAGATAACTT 613  
DB 751 rProGluThrSerAlaGluAspGluGlnGlnAsnProSerLeuSerAlaAsnLysAsnAr 771  
QY 614 TTATCAACATAAATGAAAAATAGAAAAATGTTTCTCAGCTAGGTTTGTAT----- 665  
DB 771 gValThrGluLysIleAspGluGlyGluAsnIleSerPheProGlyGlyAspAspSe 791  
QY 666 -----AAGTCACAGTTGAAAAAGGTACAGTATTGTAAGACGATCAGACTGC 715  
DB 791 rValValIleAsnSerAsnValGluGlnSerAsn-----ValGluThrGluAspAl 808  
QY 716 GCCTATGTGTAAGTGGCGAATGAGGCGCCACACTCAGAACGGCTTTTGGAAAGTGAACC 775  
DB 808 aGlyAsnSerProIleGlnAspGluValSerGlnGluGlyArgIleLeuAsnGluTh 828  
QY 776 TCCACGCGTAACTCTGTCGACAGACAGTTCAGTAATGCTTAATGTCAGTCTGCTCCC 835  
DB 828 rAspIleValAspThrValAlaLysValIleGluAsnGluLysIleSerProIleAsnSe 848  
QY 836 AAAGGATGATCAGCTGACACAAATAGTAGGAGAGTAGAGATATATCAGCAGTTTTCAC 895  
DB 848 rLeuAspAspHisThrGluLeuAlaThrAspSerGlyAsnAspSerAsnSerThrGluSe 868  
QY 896 ACATGTAAGCTTGCAGATGCAAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 955  
DB 868 rAspIleGlnSerLysAsnGluIleSerProValIleAsnGluLysAsnThrGluLeu-- 887  
QY 956 AAGCCACACAGAGTGTGGCAAGCTTGCCTCAGAGAGCTGTGTGACAGGCTGTGACAGA 1015  
DB 888 -----IleGlnLys-----HisIleGluSerIleLeuAlaAspLysArgLe 901  
QY 1016 GGAGGAG-----ACAGACGT 1030

1202 GGATTTCAGAACGATGAGACAGCAAGTCCAGGTTTGTATGAAACAGAAATAGCAGTTTC 1261  
DB 928 aGlySerLysTrpAspSerGlu-----SerAsnSe 938  
QY 1262 TGCTCAACAGCAAAATAACTTCAAGTTCACCAAGAGAGAGCTGACACTGAGTTGAG 1321  
DB 938 rGluGlnAspValThrLysSerArgLysSerAspPro-----Ar 951  
QY 1322 GAAGCGGTCTCTGCTAAGGAGGTGAGATTGCAATTTCATTTTCAATTTGAAGGAGGAGA 1381  
DB 951 gArgGlySerGluLysGluGluGlyGluAlaSer-----SerAspSerGl 966  
QY 1382 GAGTCGAGCTGGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1421  
DB 966 uSerGluValGlyGlnSerHisIleLysAlaLysProProAlaLysProProThrSerTh 986  
QY 1422 -----GGAAGTACTTCTAG 1435  
DB 986 rPheLeuProGlySerAspGlyAlaTrpLysSerArgArgProGlnSerSerAlaSerGl 1006  
QY 1436 CCTGAATGAGTGCAGAAAT-----TCTAAGCAACATGG 1471  
DB 1006 uSerGluSerSerCysSerAsnLeuGlyAsnIleArgGlyGluProGlnLysGlnLysHi 1026  
QY 1472 GAGAAAGGATTTCTAAATTCACAGATCATTTTCATGAGAGTCCCAAGAGCAGGACAAAG 1531  
DB 1026 sSerLysAspAspLeuLysGlyAspHisThrLysArg-----AlaArgGluLysSe 1043  
QY 1532 AAAGAACAAATGTAATGAAATCAACAAACAAAGACAGAGAGAGATCCCTAAATACATTC 1591  
DB 1043 rLysAlaLysLysAspLysLysHisLysAlaProLysArgLysGlnAlaPheHisTrpGl 1063  
QY 1592 ACCTCACCTTTCT-----CCAGATAGAAATGCTTGG 1624  
DB 1063 nProLeuGluPheGlyAspAspGluGluGluMetAsnGlyLysGlnValThrGl 1083  
QY 1625 AACTCTATTGAGGAGATGAGAGAGATGCCAGGTGT-----GGATCCG 1669  
DB 1083 nAspProLysGluLysArgHisValSerGluLysCysGluAlaValLysAspGlyLe-- 1102  
QY 1670 GCTGCTCCTTGAGA-----CCATCTGCCAATCACACAGT 1705  
DB 1103 ---ProAsnValGluLysThrCysAspGluGlySerSerProSerLysProLysLysGl 1121  
QY 1706 GACTATTCCGGTAGAT---CTTTTCGAATAGGAGAGTTCCTAAACCTTTCCCAACACA 1762  
DB 1121 yThrLeuGluGlnAspProLeuAlaGluGlyGlyHisAspProSerCysProAlaPr 1141  
QY 1763 TTTTAAAGATTGTGGGACACACAGCATGTTTAAGATGCTTTTTCAGACACAACTTGTGA 1822  
DB 1141 oLeuLys--ValGluAspAsnThr---AlaSerSerProProSerAlaGlnHisLeu-- 1158  
QY 1823 CCTGTGGAAGATGAGATGGT-----1844  
DB 1159 -----GluGluHisGlyProGlyGlyGluAspValLeuGlnThrAspAsAs 1175  
QY 1845 -----GAGCGAGCTGCAGCGACCGCTGGGAACTCACTTACAGCTGC 1885  
DB 1175 nMetGluIleCysThrProAspArgThrSerProAlaLysGlyGluValValSerProLe 1195  
QY 1886 ACTTCTCAACAGCTCCTCGGCCCGAC 1913  
DB 1195 uAlaAsnHisArgLeuAspSerProGlu 1204  
RESULT 13  
T03277  
pol protein - yeast (*Candida albicans*) retrotransposon pCal (fragment)  
C;Species: *Candida albicans*  
C;Date: 24-Mar-1999 #sequence\_revision 24-Mar-1999 #text\_change 29-Oct-1999  
C;Accession: T03277  
R;Matthews, G.D.; Goodwin, T.J.D.; Butler, M.I.; Berryman, T.A.; Poulter, R.T.M.  
J. Bacteriol. 179, 7118-7128, 1997



```

Db      901 uAspGluPheGluThrTyraAsnValaspGluLeuGluValIleAsnAspAspIle 921
      ...|||
QY      1031 GGTGTCCGAGACCCCTTGTCCGACACTGGCTCTGAG---GATGTTGGTACTGGACTGAA 1087
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      921 eAlaGluAlaAsnProLeuProAspGluAsnAsnAspValGlnMetAsnGluSerPheAs 941
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY      1089 AAATGCCAACAGATTGAATAGACAGAAAGTAGTCTAGGAATTTCTCCCTCCATTTGAGAA 1147
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      941 pAsnAsnHisSerMetSerArgAlaAlaLysLysLysTyra-----ThrPheGluLys 957
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY      1148 AGAAAGTGAACCTGAGTCCCAATGATGATAGATATTCAAAAAATAGTTGTGAGGATTC 1207
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      957 sGluValAsnGluLys-----IleAlaGlyThrLysHisSerLeu----- 970
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY      1208 AGAAGCAGATGAGAGACAAAGTCCAGGTTTGTATGACAGGAGATAGAGTCTGCTCA 1267
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      971 -----AspThrThrAspPro-----ArgGluAlaIleArgValle 982
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY      1268 AACAGCAATAAACCTTCAAGGTTCCACCAAGAGAGCTGAC-----ACTGA 1315
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      982 uAsnThrGlyGluThrLysArgIleGluProLysLysArgGluValProIleThrVally 1002
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY      1316 GTTGAGGAAGCGTCTCTGCTAAGGAGGTGAGATTCGATTACATTTCCAAATTTGAAGG 1375
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      1002 sIeuAsnLysArgSerGlnTyrlsSerProTyraVal----- 1014
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY      1376 AGGAGAGATCGAGCTGGAGTAATGATGTAATGCCAACGA----- 1418
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      1015 -----ThrArgSerGlyArgThrValIleAsnProLysArgTyraLysAlaValVa 1032
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY      1419 -----CCTGGAAGTACTTCTAGCTGAATGAGTGCAGAGAA 1456
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      1032 lAsnLysIleAspTyraAsnAspProGlyTrpIleLysSerMetAsnAlaGluLeuGluLys 1052
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY      1457 TTCTAAGCAACATGGGAGAAAGAT-----TC 1483
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      1052 sPheArgSer-----LysAspValTyraGluValProIleProThrGlyVally 1069
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY      1484 TAAATACAGATCATTTTCATGAGAGTGCACCAAGCAGAGGAC-----AAAG 1531
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      1069 sProIleSerMetGlyTrpValHisThrGluLysIleAspSerLeuLysGlyValValar 1089
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY      1532 AAAGACACATCTGAATGAACATCAACAGACAGAAAGAGATC----- 1577
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      1089 glySerArgCysValValHisGlyAsnArgGlnLysGluLysLysAspTyraAspProph 1109
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY      1577 ----- 1577
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      1109 eSerValSerSerProValIleAspLeuValThrIleArgLeuLeuThrIleIleGlyCy 1129
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY      1577 ----- 1577
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      1129 sGluLeuGlyMetThrIleGlnHisLeuAspValGluSerAlaTyraLysAsnAlaSerIl 1149
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY      1578 -----CCTAAATACATT-----CCACCTCACCTTCTCCAGATAAGAAA----- 1616
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      1149 eThrHisSerAsnProIleTyraValPheProLysSerValProLeuLysLysAsnHi 1169
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY      1617 -----TGCTTGGAACTCCTATTGAGAGATGAGGAGA-----ATGCCAAGGTG 1660
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      1169 sCysTrpLeu-----LeuLysArgSerValTyraGlyLeuLysGlnSe 1183
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY      1661 TGGATCCGGTGGCTCCCTTGAGACCATCTGCCAATCACACAGTACTATTCCGTA-- 1718
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      1183 rGlyLeuGluLys-----TyraHisThrIleLys---ArgValle 1195
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY      1718 ----- 1718
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      1195 uGluAspIleGlyPheThrGlnValLeuHisAsnAspGlyLeuPheHisIleGluTyraGl 1215
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY      1719 -----GATCTTTGGCAATAGAGAGAT 1741
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

```

Db      1215 uGluGlySerValIleTyraLeuGlyLeuTyraValAspAspIleLeuMetValGlySerSe 1235
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY      1742 TCCTAAACCTTTCCCAACACATTTTAAAGATTGTGGGACCAACAGCATGTTTAAGATGCC 1801
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      1235 rGlnLysValIleAspAsnPheValAspGlnLeuArgAspHisPheGluValLysValPh 1255
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY      1802 TTGTTTCAGAACAAAATTTGTACCTGTGTGAGATGAGAAATGTGTGAGCGAGCTGCAGGAG 1861
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      1255 eGlyGluIleSerAsnTyraLeuGlyIleGluPheArgLysThrGlu-----Se 1271
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY      1862 CCGGTGGGAACCTCATTCAGATGCTCTCAACAGGCTCACTCGCCCCCAGAACCTGAA 1921
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      1271 rGlyTyraIleLeuSerGlnGluLysPheLeuLysLysLeuLysAspPheLysLeuAs 1291
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY      1922 GGATGCTATTCTGAAGTACAATGTGGCATAT-----TCTAAGAATATGGAGCTTTTACAC 1975
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      1291 pAspSerTyraGlyLys---AsnIleProTrpIleProAsnAspLysTyraGluLysValAl 1310
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY      1976 TTTGATT-----CATTTCTGGGATAAGGTA----- 2000
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      1310 aIleIleArgGluAsnValAsnProGluAsnAspPhe---GluLysValProAsnGluTh 1329
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY      2001 -CTAGAGAGAGCAGAGAGCTCAACACTTGTATCAGTCC----- 2036
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      1329 rLeuLeuAspProAspAlaLysLysLeuTyraGlnSerGlyValGlySerLeuLeuTrpAl 1349
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY      2037 -----ATCTTGCCCTGATATGTTGAAATTCACCTCTGTCTGCCAAATATTGTGTAC 2086
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      1349 aAlaThrAsnThrArgProAspIleSerValValAsnSerLeuGlySerLysSerAl 1369
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY      2087 CCAGCCA-----ATACCACTCTGTAACAGAA 2113
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      1369 aAsnProAsnValHisAspTyraGluLysLeuIleTyraCysLeuArgTyraLysLeuAsnSe 1389
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY      2114 GATGATCATTCATCCATCAGATGTCACAGGAACAGATTGCCAGTC 2157
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      1389 rMetGlyTyraHisIle-GluTyraLysArgAsnArgLeuAsnIle 1403
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

## RESULT 14

T29757 protein UNC-99 - Caenorhabditis elegans

C;Species: Caenorhabditis elegans

C;Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 03-Dec-1999

C;Accession: T29757

R;Du, Z.; Le, T.T.; Wilson, R.

submitted to the EMBL Data Library, May 1997

A;Description: The sequence of C. elegans cosmid C09D1.

A;Reference number: Z20679

A;Accession: T29757

A;Status: preliminary; translated from GB/EMBL/DBD

A;Molecule type: DNA

A;Residues: 1-6642 &lt;DUZ&gt;

A;Cross-references: EMBL:AF003131; PIDN:AA54132.1; GSPDB:GN00019; CESP:unc-89

A;Experimental source: strain Bristol N2; clone C09D1

C;Genetics:

A;Gene: CESP:unc-89

A;Map position: 1

A;Introns: 17/2; 108/3; 154/2; 211/2; 265/3; 326/2; 352/3; 426/2; 454/1; 500/1; 537/1; 61

/3; 5917/1; 6027/1; 6061/3; 6153/2; 6515/1; 6552/3; 6609/1

## Alignment Scores:

Pred. No.: 0.00196 Length: 6642

Score: 170.50 Matches: 198

Percent Similarity: 35.12% Conservative: 142

Best Local Similarity: 20.45% Mismatches: 418

Query Match: 2.35% Indels: 210

DB: 2 Gaps: 47

US-09-302-812-1 (1-4070) x T29757 (1-6642)

QY 246 GCGGGTCCCGCATGAGTGGCGGCCCGCGGTGTGAG---CCCTGCACCAAGCGACCCCGC 302

Db 1394 AlaSerProThrLysLysThrGlyGluValLysSerProLysGluLysSerProAla 1413



QY	303	TGGAGCGCGCTGCAACTTCTCGCGCGCGCTCGAGCGCCGAGAGTCTCCCGCGCAGG	362
Db	1414	SerProThrLysLysGluLysSerProAlaAlaGluGluValLysSer---ProThrLys	1432
QY	363	CAGAGCGCGCTCCTGATTCCAAGGACGCTCGGTGTCAGTTCAGGGTCCCGCGCTCTCG	422
Db	1433	LySGluLys-----SerProSerSerProThrLysLysGluLysSerProSerSer	1449
QY	423	TCAGGTCGCGCTGGCGCGGCGGACAGCCGAGGCGCGCCACTCTCTGTGTTTC	482
Db	1450	ProThrLysLysThrGlyAspGluValLysGluLysSerProProLysSerProThrLys	1469
QY	483	AAACAGAGAGACTATAACACGATGGATGGACATAAA-----	518
Db	1470	LysGluLysSerProGluLysProGluAspValLysSerProValLysLysGluLysSer	1489
QY	519	-----GGAATCAACAGACITGAATCAGAA---AGTTTGCATAGTAAGAAGAAC	563
Db	1490	ProAspAlaThrAsnIleValGluValSerSerGluThrThrIleGluLysThrGluThr	1509
QY	564	AACAATCAAGAGAA-----GAATCCATGATGATGTTCTGTACAAAAGAT	608
Db	1510	ThrMetThrThrGluMetThrHisGluSerGluLysSerArgThrSerValLysLysGlu	1529
QY	609	AACTTTTATCAACATAACATCGAATAAATAGAAATGTTTCTCAGCTAGGTTTGTATAG	668
Db	1530	LysThrProGluLysValAspGluLysProLysSerProThrLysLys-----AspLys	1547
QY	669	TCACGAGTTGAAAAGGTACACAGTATTCAAGCAGCATCAGACTCGGCTATGCTGAAG	728
Db	1548	SerProGluLysSerIleThrGluGluIleLysSerProValLys-----	1562
QY	729	TGGCAGAAATGAAGCGCCACACTCAGAACGGGTTTTGGAAAAGTGAACCTCCAGCGGTA	788
Db	1563	---LysGluLysSerProGluLysValGluGluLysProAlaSerProThrLysLysGlu	1581
QY	789	CTGTTACACAGCAGTTCAGTAATGCTAATGTCATCAGTCGTCCTCCCAAGAGTATCAC	848
Db	1582	LysSerProGluLys-----ProAlaSerProThrLysLysSerGluAsnGluValLys	1599
QY	849	AGTGACACAATAGTGGAGCAGTAGAGATAATCAGCAGTTTTTTGACACATGTAAAGCTT	908
Db	1600	SerProThrLysLysGluLysSerProGlu---LysSerValValGluGluLeuLysSer	1618
QY	909	GCGAATGCAAAAGCAGCAGATGGAA---GATGAACAGGCGACAGAACGCCAAGCCAC	965
Db	1619	ProLysGluLysSerProGluLysAlaAspAspLysProLysSerProThrLysLysGlu	1638
QY	966	AAGTGTGCGAAGGCTGCCATCTTCGACAGCCTGTGCAGGCTGTCCAGCAGGAGGAGACA	1025
Db	1639	LysSerProGluLysSerAlaThrGluAspValLysSerProThrLysLysGluLysSer	1658
QY	1026	---GACGTGGTGTCCGAGAGCCCTTGTTCGACACTGGCTCTGAGGATGTGGTACT---	1079
Db	1659	ProGluLysValGluGluLysProThrSerProThrLysLysGluSerSerProThrLys	1678
QY	1080	-----GGACTGAAAATGCCAACAGATGTAATAGACAAGAAAGTAGTCTAGGA	1127
Db	1679	LysThrAspAspGluValLysSerProThrLysLysGluLysSerProGluThrValGlu	1698
QY	1128	AATTCTCCT-----CCATTTGAGAAAGAAAGTGAACCTGAGTCACCAATGATGTAGAT	1181
Db	1699	GluLysProAlaSerProThrLysLysGluLysSerProGlu-----	1712
QY	1182	AATTCACAAAATAGTTGTCAGGATTCAAGACAGATGAAGAGACAAGTCCAGGTTTGTAT	1241
Db	1713	-----LysSerValValGluGluValLysSerProLysGluLysSerProGluLysAla	1730
QY	1242	GAACAGGAGATAGCAGTTCTCCTCAACAGCAATAAACCCTTCAAGGTTCCACCAAGA	1301
Db	1731	GluGluLysProLysSerProThrLysLysGluLysSerProGluLysSerAlaAlaGlu	1750

QY	1302	GAAGCTGACACTGAGTTGAGGAAG---CGGTCTCTGCTAAGGAGGAGTGCAGATTGCATTA	1350
Db	1751	GlUVallysSerProThrIlysLysGluLysSerProGluLysSerAlaGluGluLysPro	1770
QY	1359	CATTTCCAATTTGAAGGAGGAGGAGCTCGAGCTGGAATGAATCATGCTGAATGCCAAACGA	1418
Db	1771	LysSerProThrIlysLysGluSerSerProValLysMetAlaAspAspGluValLysSer	1790
QY	1419	CCT-----CGAAGTACTTCTAGCCGTAATGTATAGATGCGAATAATCTTAAGCAA	1466
Db	1791	ProThrLysLysGluLysSerProGluLysValGluGluLysProAlaSerProThrLys	1810
QY	1467	CATGGGAGAAAGATTCTTAATCATCAGATCATTTTCATGAGAGTGCC-----	1514
Db	1811	LysGluLysThrProGluLysSerAlaAlaGluGluLeuLysSerProThrLysLysGlu	1830
QY	1515	-----AAAGCAGAGGACAAAGAAAGAAACAAATGT---GAA	1547
Db	1831	LysSerProSerProThrLysLysThrGlyAspGluSerLysGluLysSerProGlu	1850
QY	1548	ATGAAACATCAAGAACACAGAAAGAGATCCCTAAATACATATCCACCTCACCTTCTCCA	1607
Db	1851	LysProGluLysLysProLysSerProThrProLysLysSerProGly---SerPro	1869
QY	1608	GATAAGAAATGGCTTGGAATCTCTATTTAGGAGATGAGGAGATGCCAAGTGTGGGATC	1667
Db	1870	LysLysLysLysSerLysSerPro-----GluAlaGluLysPro-----	1883
QY	1668	CGGCTGCCTCCCTTGAGACCATCTGCCAATCACACAGTGCATCTTCGGGTAGATCTTTTG	1727
Db	1884	---AlaProLysLeuThrArgAspLeuLysLeuGlnThrValAsnLysThrAspLeuAla	1902
QY	1728	CGAATAGGAGAGTTCTTAAACCTTTTCCCAACACATTTTAAAGATTTGTGGGACACACAAG	1787
Db	1903	HisPhe---GluValValValGluHisAlaThrGluCysLysTrpPheLeuAspGlyLys	1921
QY	1788	CATGTTAAGATGCTTGTTCAGAACAAACTGTACCTCTGGAAGTACAGATGGTGAG	1847
Db	1922	GluIle-----ThrThrAlaGlnGlyValThrValSerLysAspAspGlnPheGlu	1938
QY	1847	-----	1847
Db	1939	PheArgCysSerIleAspThrThrMetPheGlySerGlyThrValSerValValAlaSer	1958
QY	1848	CGAGCTCGAGGACGCGGTGGGAACCTCATTCACAGACTGCACATT---CTCACACGGCTCACT	1904
Db	1959	AsnAlaAlaGlySer-----ValGluThrLysThrGluLeuLysValLeuGlu	1974
QY	1905	CGGCCCCAGAACCTTGAAGATGCTATTTCTGAAGTACAATGTGGCAATTCTTGAAGATGG	1964
Db	1975	ThrProLysGlu-----ThrLysLysPro-----	1982
QY	1965	GACTTTACA---GCTTTGATGATTTC-----TGGGATAGGTACTAGAGAA	2009
Db	1983	GluPheThrAspLysLeuArgAspMetGluValThrLysGlyAspThrValGlnMetAsp	2002
QY	2010	GCAGAAGCTCAACAC-----TTGTATCAG-----	2033
Db	2003	ValIleAlaLeuHisSerProLeuTyrlsTrpTrpGlnAsnGlyAsnLeuLeuGluAsp	2022
QY	2034	-----TCCATCTCCCTGATATGTGTGAATTCACCTCTCTCTGCCAAAT	2078
Db	2023	GlyLysAsnGlyValThrIleLysAsnGluLysSerSerLeuIleProAsn	2042
QY	2079	ATTGTACCCAGCCAAATACCACCTCTGAAACAGAGATGAATCATTCATCAACAATGCA	2138
Db	2043	-----AlaGlnAspSerGlyLysIleThrValGluAlaSerAsnGluValGly---Ser	2059
QY	2139	CAGGAACAGATTGCGAGTCTTTTAGCTTAATGCTTTCTTCTCGACG-----	2183
Db	2060	SerGluSerSerAlaGlnLeuThrValAsnProProSerThrThrProIleValValAsp	2079
QY	2184	TTTCCACGA-----CGCAATGCCAAGATGAATCGAGTATTCC	2222



```

Db 588 AspAspTyrAspArgAspSerGluSerGluGluAspAsnLeuGlyArgLeuSer----- 605
QY 903 AAGCTTGCAGATCAAGACAGCATGAGATGACACAGCGCAGAGCCAGAGCCAC 962
Db 606 GlyArgAlaGluGlyLysSerLysLeuThrAlaGlnLysSerHisMetSerGluGlyPro 625
QY 963 CAGAGTGTGGCAGGCTTCCATCCTGCA-----GAAAGCTGTGCAGGGTGT 1010
Db 626 AspAspLeuGlyArgTyrPhePheProSerAspThrGluAspGlnGlyAspAspSerLys 645
QY 1011 CAGCAGGAGGAGCAGACGCTGTGTCGAGAGCCCTTGTGCGACACTGGCTCTGAG--- 1067
Db 646 ThrGlnGluGluSerAsp-----AlaGluThrPro-----ThrGlyLeuLysPhe 660
QY 1068 ---GATGTTGTACTGACTCAAAATGCCAACAGATTGAATAGACAAAGTAGTCTA 1124
Db 661 GlyProLeuAlaSerGlyLeuGluAsnGluThrThrLeu-----ProSerTyr 676
QY 1125 GGAATTTCTCTCCATTTGAGAAAGAAAGTGAACCTGAG-----TCACCAATG 1172
Db 677 GlySerSerProProArgAspLysThrSerSerLysSerIleLysGluTyrLeuProThr 696
QY 1173 GATGTAGATAATTCAAAATAGTTGTGAGGATTCAAGACAGATGAGAGACAAAGTCCA 1232
Db 697 GluValAspProSerArgSerSerSerLeuGlnThrAlaSerSerSerSerIleArgAsn 716
QY 1233 GCTTTTATGACAGGAGATAGCAGTCTGCTCAACAGCAAAATAACCTTCAAGGTTTC 1292
Db 717 GluLeuTyrThrGlnLysAlaSerAsnSerAspLys-----ArgProSerSerIle 733
QY 1293 CAACCA-----AGAAAGCTGACACTGAGTTGAGGAGCGGTCTCT 1334
Db 734 ProProAspSerSerSerSerAspAspGluSerAspMetGluLeuProLysArgValSer 753
QY 1335 GCTAAGGAGGTGAGATTTCGATTACATTCCAAATTTGAAGGAGGAGAGTCGAGCTGGA 1394
Db 754 PheArgTyrGlnGluLysArgThrGluSerArgThrArgProThrHisLeuHisSerGly 773
QY 1395 ATGAAT-----GATGTAATGCCAACGACGCTGGAAGTACTTCTAGCCTGAATGTAGAG 1448
Db 774 ValSerHisLysAspLeuGluGluLeuProThrArgAlaSerThrArgSerGlnAsp 793
QY 1449 TGCAGAAATTTCTAAGCAATGGGAGAAAGGATTCTAAATCAGAGATCATTTCATGTAGA 1508
Db 794 ArgArgThrHisLysThrThrProAlaSerAlaSerTyrPheHisThrMetSer 813
QY 1509 GTGCCAAAGCAGGACAAAGAAAGAACACATGTGAATGAAACATCAAGAACAA--- 1565
Db 814 ---SerAspAspGluAspGluLysGluValHisArgAspThrAlaHisIleGlnThrArg 832
QY 1566 -----GAAAGG----- 1571
Db 833 ProTyrIleSerIleSerArgArgThrLysGlyGlnGluArgArgProSerLeuValThr 852
QY 1572 ---AAGATCCCTAAATAC-----ATTCACCTCTCACCTTTCTCCAGATAAG 1613
Db 853 AlaLysIleAspLysValSerPheAspGluGluSerProProLysLeuSerProGluAla 872
QY 1614 AAA 1616
Db 873 Lys 873

```

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: May 26, 2004, 15:32:43 ; Search time 34.05 Seconds  
(without alignments)  
12447.891 Million cell updates/sec

Title: US-09-302-812-1

Perfect score: 7242

Sequence: 1 accggaagtgaacgaagcc.....aaattttcattcaacaaaaa 4070

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 283362

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame\_n2p.model -DEV=xlp  
-Q=/cgn2.1/USPTO spool\_p/63331148/runat\_26052004.150053\_5664/app.query.fasta\_1.12437  
-DB=SwissProt 42 -QFMT=fastan -SUFFIX=n2p.rsp -MINMATCH=0.1 -LOOPCL=0  
-LOOPTXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOALIGN=200 -THR SCORE=ptc -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=63331148 @CGN 1.1.128 @runat\_26052004.150053\_5664 -NCPU=6 -ICPU=3  
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : SwissProt\_42.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	184	2.5	3664	1 MINT_HUMAN	Q96t58 homo sapien
2	180.5	2.5	1447	1 BUD4_YEAST	P47136 saccharomyc
3	174.5	2.4	1462	1 NKCR_HUMAN	P30414 homo sapien
4	174	2.4	1359	1 ATRX_CAEEL	Q9u7e0 caenorhabdi
5	173	2.4	1210	1 AF4_HUMAN	P51825 homo sapien
6	171.5	2.4	1453	1 NKCR_MOUSE	P30415 mus muscucu
7	171	2.4	5147	1 PCLO_HUMAN	Q9y6v0 homo sapien
8	170.5	2.4	6632	1 UN89_CAEEL	O01761 caenorhabdi
9	169.5	2.3	913	1 IP38_HUMAN	Q99613 homo sapien
10	168	2.3	1073	1 PVDA_PLAKN	P22545 plasmodium
11	164.5	2.3	646	1 SG1_BOVIN	P23389 bos taurus
12	163	2.3	1070	1 PVDC_PLAKN	P50494 plasmodium
13	163	2.3	1362	1 BRD4_HUMAN	O60885 homo sapien
14	162	2.2	513	1 DMP1_HUMAN	Q13316 homo sapien
15	162	2.2	2158	1 MY9B_HUMAN	Q13459 homo sapien
16	161.5	2.2	943	1 YM11_YEAST	P39523 saccharomyc
17	161.5	2.2	5560	1 SPEN_DROME	Q88x83 drosophila
18	160	2.2	633	1 MLH_TETTH	P40631 tetrahymena

19	158.5	2.2	5596	1 MDN1_HUMAN	Q9nu22 homo sapien
20	158	2.2	1424	1 NCO3_HUMAN	Q9y6q9 h nuclear r
21	158	2.2	3644	1 MINT_MOUSE	Q62504 mus muscucu
22	157	2.2	1411	1 TCOP_HUMAN	Q13428 homo sapien
23	157	2.2	3122	1 DPOZ_MOUSE	Q61493 mus muscucu
24	156.5	2.2	773	1 ELAI_HUMAN	Q8cb77 mus muscucu
25	156	2.2	1341	1 ACIN_MOUSE	Q9ukv3 homo sapien
26	155	2.1	489	1 DMP1_RAT	P98193 rattus norv
27	154.5	2.1	1338	1 ACIN_MOUSE	Q9jix8 mus muscucu
28	154	2.1	406	1 SR40_YEAST	P32583 saccharomyc
29	153.5	2.1	5085	1 PCLO_RAT	Q9jks6 rattus norv
30	153	2.1	5120	1 PCLO_CHICK	Q9p36 gallus gall
31	152.5	2.1	472	1 YWIE_CAEEL	Q23525 caenorhabdi
32	152.5	2.1	915	1 CE05_HUMAN	Q9nyf5 homo sapien
33	152	2.1	914	1 MYS3_YEAST	P54784 saccharomyc
34	152	2.1	2004	1 ORC1_HUMAN	Q92794 homo sapien
35	151	2.1	759	1 EPLI_HUMAN	Q9ubh6 homo sapien
36	151	2.1	1311	1 ATRX_DROME	Q99qns drosophila
37	150.5	2.1	503	1 DMP1_MOUSE	O55188 mus muscucu
38	150.5	2.1	1615	1 RIM1_RAT	Q9jjr4 rattus norv
39	150.5	2.1	2224	1 FAS_HUMAN	P12259 homo sapien
40	150.5	2.1	5038	1 PCLO_MOUSE	Q9gyx7 mus muscucu
41	150	2.1	771	1 CALD_CHICK	P12357 gallus gall
42	149.5	2.1	1654	1 PCFB_HUMAN	Q94913 homo sapien
43	149.5	2.1	1692	1 RIM1_HUMAN	Q86ur5 homo sapien
44	149.5	2.1	2453	1 NCR1_MOUSE	Q60974 mus muscucu
45	149	2.1	845	1 NFM_RAT	P12839 rattus norv

#### ALIGNMENTS

RESULT 1  
MINT\_HUMAN  
ID MINT\_HUMAN STANDARD; PRT: 3664 AA.  
AC Q96t58; Q9y6q9; Q9NWH5; Q9UQ01; Q9Y556;  
DT 10-OCT-2003 (Rel. 42, Created)  
DT 10-OCT-2003 (Rel. 42, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Mx2-interacting protein (SMART/HDAC1 associated repressor protein).  
GN MINT OR SHARP OR KIAA0929.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A., FUNCTION, SUBCELLULAR LOCATION, INDUCTION,  
RP RNA-BINDING, AND INTERACTION WITH NCR2; HDAC1; HDAC2; RBBP4; MED3;  
RP RAR AND MTALL1.  
RC TISSUE=Liver, and Pituitary;  
RX MEDLINE=21231190; PubMed=11331609;  
RA Shi Y., Downes M., Xie W., Kao H.-Y., Ordentlich P., Tsai C.-C.,  
RA Hon M., Evans R.M.;  
RT "Sharp, an inducible cofactor that integrates nuclear receptor  
RT repression and activation.";  
RL Genes Dev. 15:1140-1151(2001).  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Bird C.;  
RL Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE OF 294-3664 FROM N.A.  
RA Rhodes S., Huckle E.;  
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.  
RN [4]  
RP SEQUENCE OF 793-1595 FROM N.A., AND VARIANT PRO-1091.  
RC TISSUE=Embryo, and Teratocarcinoma;  
RA Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,  
RA Nishikawa T., Nagai K., Sugano S., Takahashi-Fujii A., Hara H.,  
RA Tanase T., Nomura Y., Togiya S., Komai F., Hara R., Takeuchi K.,  
RA Arita M., Nabekura T., Ishii S., Kawai Y., Saito K., Yamamoto J.,  
RA Wakamatsu A., Nakamura Y., Nagahari K., Masuho Y., Oshima A.;  
RT "NED0 human cDNA sequencing project.";  
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.



QY 207 TGAGCTCTACGCCG-----GGCTGGCCCGGAGACTCAGT 242  
 Db 1966 nGluAlaLysGluProAlaGluThrLeuLysProProGluGlyTrp--ArgSerProArg 1985  
 QY 243 GCTGGGCTCCAGCATGAGTGGCGGCGCGCTGTGAGCCCTGCACCAAGGACCCCGC 302  
 Db 1986 SerGlnLysThrAlaAlaGlyGlyProGlnGlyLysLysGlyLysAsnGluProLys 2005  
 QY 303 TGGGAGCGCTGCACTCTCCGCGCGCGCTCGGAGCGCGGAGCTTCCCGGCGAGG 362  
 Db 2006 ValAsp-----AlaThrArgProGluAlaThrThrGluVal-----GlyPro 2019  
 QY 363 CAGAGCGCGCTCCTCATTCAGAGACGCTCGGTCGACATTCAGGCTCCCGCGCTCG 422  
 Db 2020 GlnIleGlyValLysGluSerMetGluPro-----LysAlaAlaGluGlu 2036  
 QY 423 TCAGCTGCGCTCGGCGCGCGGAGCAGACCGAGCGGCGGAGCCCTCTCTTTTTC 482  
 Db 2037 AlaGlySerGluGlnLysArgAspArgLysAspAlaGly----- 2049  
 QY 483 AAACAGAGACTATAACCATGGATGGACACTAAAGGAATCAAGACAGTTCAATCAGAA 542  
 Db 2050 -----ThrAspLysAsnProProGluThrAlaProValGluValValGluLysLys 2066  
 QY 543 AGTTTGCATAGTAAGAAACAAACATACAGAGAGAAATCCATGATGATGATCTGTACAA 602  
 Db 2067 ProAlaProGluLysAsnSerLysSerLysArgGlyArgSerArgAsnSerArgLeuAla 2086  
 QY 603 AAAGATACTTTTATCAACATACATGAGAAATAGAAATTTTCTCAGTAGTTT 662  
 Db 2087 ValAsp-----LysSerAlaSerLeuLysAsnValAspAlaAlaValSer 2101  
 QY 663 GATAGTACACAGTTCAAAAGGTACACAGTATTTGAAGCAGCATCAGAGCTGGGTATG 722  
 Db 2102 ProArgGlyAlaAlaAlaGlnAlaGlyGlu-----ArgGluSerGlyValValAlaVal 2119  
 QY 723 TGTAAGTGCAGAAATCAAGGCGCACACTCAGAA-----CGGCTTTTGGAAAGT 770  
 Db 2120 SerProGluLysSerGluSerProGlnLysGluAspGlyLeuSerSerGlnLeuLysSer 2139  
 QY 771 GAACCTCCAGCGTAACTCTGTGTCAGAGAGCTCAGTAACTGCTAACTGATCAGTACG 830  
 Db 2140 AspProValAspProAspLysGluProGluLys-----GluAspValSerAlaSer 2156  
 QY 831 TCCCCAAAGATGATCAGCTGACACAAATAGTAGAGAGTAGATGATGATCAGCAGTTT 890  
 Db 2157 GlyProSerProGlu-----AlaThrGlnLeuAlaLysGlnMetGluLeuGlnAla 2174  
 QY 891 TTGACACATGTA---AAGCTTGGGAATGCAAGCAGCAGATGGAAGATGAACAGGCGAGA 947  
 Db 2175 ValGluHisIleAlaLysLeuAlaGluAlaSer----- 2185  
 QY 948 GAAGCCAGAGCCACAGAAAGTGTGGAAGCTTGCATCTCCTCAGAGAGCTGTGCGAGG 1007  
 Db 2186 -----AlaSerAlaAlaLysLysAlaAspAlaProGluGly 2197  
 QY 1008 TGTCAGCAGGAGGAGCAGAGCTGGTGTCCGAGAGCCCTGTGCGGACATGCTGTGAG 1067  
 Db 2198 LeuAlaProGluAspArgAsp-----LysProAlaHisGlnAlaSerGluThr 2213  
 QY 1068 GATGTTGGTACTGACTGAAATAAT---GCCAACAGATTTGAATAGACAGAAAGTAGTCTA 1124  
 Db 2214 GluLeuAlaAlaIleGlySerIleAsnAspIleSerGlyGluProGluAsnPhe 2233  
 QY 1125 GGAATTTCTCTCATTTGAAAGAAAGTGA----- 1157  
 Db 2234 ProAlaProProProTyProGlyGluSerGlnThrAspLeuGlnProProAlaGlyAla 2253  
 QY 1158 -----CCTGATCACCATGATGATGATGATGATGATGATGATGATGATGATGAT 1202  
 Db 2254 GlnAlaLeuGlnProSerGluGluGlyMetGluThrAspGluAlaValSerGlyIleLeu 2273

QY 1203 GATTCAGACGATGAGAGACAGTCCAGTTTGTATGATGACAGGAGATAGCAGTTCT 1262  
 Db 2274 GluThrGluAlaAlaThrGluSerSerArgProProValAsnAlaProAspProSerAla 2293  
 QY 1263 GCTCAACAGCAAAATAAACCTTCAAGGTTCCAAACCAAGAGAGAGCTGCACACTGAGTTGAGG 1322  
 Db 2294 GlyProThrAspThrLysGluAlaArgGlyAsnSerSerGluThrSerHisSerValPro 2313  
 QY 1323 AAGCGTCTCTGTAAGGAGGAGTTCGATTCGATTCATATTCATATTCATATTCATATTC 1376  
 Db 2314 GluAlaLysGlySerLysGluValGluValThrLeuValArgLysAspLysGlyArgGln 2333  
 QY 1376 ----- 1376  
 Db 2334 LysThrThrArgSerArgLysArgAsnThrAsnLysLysValValAlaProValGlu 2353  
 QY 1377 -----GGAGAGAGTTCGAGCTGGAATGAATGAATGATG 1406  
 Db 2354 SerHisValProGluSerAsnGlnAlaGlnGlyGluSerProAlaAlaAsnGluGlyThr 2373  
 QY 1407 AATGCCAAACGAGCTGGAAGTACTTCTAGCTGGAATGTAGTGCAGAAATTTCAAGCAA 1466  
 Db 2374 ThrValGlnHisProGluAlaProGlnGlu-----GluLysGlnSerGluLysPro 2390  
 QY 1467 CATGGGAGAGGATCTAAATCATCAGATCATTTTCATGAGAGTGCCTCAAGAGCAGAGAC 1526  
 Db 2391 HisSerThrProProGlnSerCysThrSerAspLeuSerLysIleProSerThrGluAsn 2410  
 QY 1527 AAAAGAAAGAAATGTAATGAATGAATCAACAAAGAGAGAGAGAGAGAGAGAGAGAG 1586  
 Db 2411 SerSerGlnGluLysSerValGluGluArgThrProThrLysAlaSerValProAsp 2430  
 QY 1587 ATTCCACTCCTCCTTCTCCAGATGAAGATGCTGGAATCTTATTCAGGAGATGAGG 1646  
 Db 2431 LeuProProProGlnPro-----AlaProValAspGlu----- 2442  
 QY 1647 AGAATGCCAGGTGTGGATCCGGCTG-----CCT 1676  
 Db 2443 ---GluProGlnAlaArgPheArgValHisSerIleGluSerAspProValThrPro 2461  
 QY 1677 CCTTTCAGAGACCATCT 1691  
 Db 2462 ProSerAspProSer 2466  
 RESULT 2  
 BUD4 YEAST  
 ID BUD4 YEAST STANDARD; PRT; 1447 AA.  
 AC P47136;  
 DT 01-FEB-1996 (Rel. 33, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 15-MAR-2004 (Rel. 43, Last annotation update)  
 DE Bud site selection protein BUD4.  
 GN BUD4 OR YJR092W OR J1905.  
 OS Saccharomyces cerevisiae (Baker's yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
 OC NCBI\_TaxID=4932;  
 RP [1]  
 RN SEQUENCE FROM N.A.  
 RA Ramezani Rad M., Kirchthath L., Hollenberg C.P.;  
 RA Submitted (SEP-1995) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: Co-assembles with BUD3 at bud sites. BUD4 and BUD3 may  
 CC cooperate to recognize a spatial landmark (the neck filaments)  
 CC during mitosis and they subsequently become a landmark for  
 CC establishing the axial budding pattern in G1.  
 CC -1- SIMILARITY: Contains 1 PH domain.



Db 917 ThrGluProArgThrGlyLeuGlylleGlyMetLeuLysThrProValLysAspValSer 936  
1580  
Db 937 IleAlaLeuAlaSerIleLysGlyTyrGluAlaSerPheSerAspThrAspSerArg 956  
1580  
Db 957 ProGluGlyMetAsnAsnSerAspAlaIleThrLeuAsnMetPheAspAspPheGluGlu 976  
1581 ---AAATACATCCCACT---CACCTTCTCCAGATAAGAAATGGCTTGA 1625  
977 AspLysMetThrProSerThrProValArgSerIleSerProIleLysArgHisValSer 996  
1626 ACTCCT---1631  
977 SerProPheLysValLysAlaGlyAsnLysGlnGluAsnGluIleAsnIleLys 1016  
1632 ---ATTGAGAGATGAGGAGATGCCAAGGTGGATCCGG---CTG 1673  
1017 AlaGluGluGluIleGluProMetThrGlnGlnGluThrAspGlyLeuLysGlnAspIle 1036  
1674 CCTCCCTTGAGCACTGCCAATCACACAGTACTATTCCGGGTAGATCTTTTGGCAATA 1733  
1037 ProProLeuLeuAlaGlnThrLysAspAsnValGluAlaLysGluGluThrIleThrGln 1056  
1734 GGAGAAGTTCCTAAACCTTTCCCAACACATTTAAAGAT---1772  
1057 LeuGluGluProGlnAspValGluGlnGluPheProAspMetGlyThrLeuTyrLeuSer 1076  
1773 ---TTGTGGGACACAAAGCATGTTAAGATGCTTCTTCA 1808  
1077 IleLysAlaIleSerThrLeuAlaLeuTyrGlyThrLys---Ser 1090  
1809 GAACAAACTTGACCTGTGGAGATGAGAAATGGTGGAGCTGCGAGCTGCGAGCAGCCGCTGG 1868  
1091 HisArgAlaThrTyrAlaIleValPheAspAsnGlyGluAsnValValGlnThrProTyr 1110  
1869 GAACACT---1877  
1111 GluSerLeuProTyrAspGlyAsnIleArgIleAsnLysGluPheGluLeuProIleAsp 1130  
1878 ---CAGACTGCATCTTCAACAGCTCACTCGGCCAGACCTGAGAGATGCT 1928  
1131 PheLysGlyLysAlaGluThrSerSerAlaSerSerGluArgAspSerTyrLysLysCys 1150  
1929 ATTCTGAAGTACAAATGGCATATCTAAG---AAATGGGACTTTACAGCTTTGATGAT 1985  
1151 ValIleThrLeuLysCysLysTyrGluLysProArgHisGluLeuValGluIleValAsp 1170  
1986 ---TTCTGGGATAGGTA 2000  
1171 LysValProValGlyLysSerPhePheGlyLysThrLysTyrLysPheGluLysLysTyr 1190  
2001 CTAGAAGAACAGAGCTCAACACTTCTATCAGTCCATCTTGGCTGATATGCTGAAAT 2060  
1191 ValGlnLysLysProLysGlnAspGluTyrAspTyrLeuPheAlaGlnAspGlySerPhe 1210  
2061 GCATCTGTCTGCCAAATATTTGTACCAGCCCAATACCACTCTCTGAAACAGAGATGAAT 2120  
1211 AlaArgCys---1213  
2121 CATTCATCAATGTCACAGAAACAGATGCGCAGCTTTTAGTAATGCTTCTTCTG 2180  
1214 ---GluIleGluIleAsnGluGlu---PheLeuLysAsnValAlaPheAsn 1228  
2181 ACCTTTCACAGCCCAATGCCAAGATCAATGATATCCAGTATTCAGAT---2234  
1229 Thr---SerHisMetHisTyrAsnMetIleAsnLysTyrSerArgIleAlaAspLysIle 1247  
2235 ATTAACCTCAATCGTGTGTGAGGAGCGTTTCATCAAGGAAACAGAGAGCTTAAACG 2294

Db 1248 HisGlySerLysArgLeuTyrGlu---LeuProArgLysAlaProHisLysValAlaSer 1266  
2295 CTC---TTCTGCTACTTTAGAAAGAGTCACAGAGAAAAAACCCTACTGGTGGTG 2345  
1267 LeuAspValGluAlaCysPheLeuGluArgThrSer---1278  
2346 ACATTCAACAAGACAGAGCTTTGAAGATTTTCCA---GAGTGGGAAGAGTGTGAAAATCTC 2402  
1279 ---AlaPheGluGlnPheProLysGlnPheSerLeuValAsnLysIle 1293  
2403 CTGACTCAGCTG---CATGTCACTTACGAGGTACCATAGAGAAACGCG 2450  
1294 ValSerLysTyrLysLeuGlnGlnAsnIleTyrLysGluGlyTyrLeuGlnAspGly 1313  
2451 CAGGCGCTGCTACAGCTGATTTTGCACACCTTTTCGTT---GGAGTGGTGTA 2501  
1314 ---GlyAspLeuGlyLysIleGluAsnArgPhePheLysLeuHisGlySerGlnLeu 1332  
2502 ACCAGTGCAGGACTTTGTGCAAGAAGAAATCCGCTTTTATCAACCTGAGTTGATTGT 2561  
1333 SerGlyTyrHisGluIleSerArgLysAlaLysIleAspIleAsnLeuLysValThr 1352  
2562 ---TCACGGCTCTTCACTGAG 2579  
1353 LysValLeuArgAsnGluAspIleGlnAlaAspAsnGlyGlyGlnArgAsnPheThrAsp 1372  
2580 GTGCTGGATCAATGAATGCTTATCATCACAGGTACTGAGCAGTACAGTGAATACACA 2639  
1373 TrpValLeuPheAsnGluCysPheGlnLeuValPheAspAsp---1386  
2640 GGTATGCGCAACATACCGCTGGCGCCGAGCCATGAAGACAGAGCGAAAGGAGCAG 2699  
1387 GlyGluArgIleThrPheAsnAlaGluCysSerAsnGluLysSer---Asp 1403  
2700 TGGCAGAGCGCAGCTGAGATGCTGCGCATCGACGCCCTCCAC---TTCAGA 2750  
1404 TrpTyrAsnLysLeuGlnGluValValGluLeuAsnValPheHisGlnProTyrValLys 1423  
2751 CGCTACTCGACAGTTTGTGCCGAGAAAGATCAGACGG 2789  
1424 LysTyrCysGluLysLeuAlaGluGluLysThrArg 1436  
RESULT 3  
ID NKCR HUMAN STANDARD; PRT; 1462 AA.  
AC P30414; 1993 (Rel. 25, Created)  
DT 01-APR-1993 (Rel. 39, Last sequence update)  
DT 30-MAY-2000 (Rel. 41, Last annotation update)  
DE NK-tumor recognition protein (Natural-killer cells cyclophilin-related protein) (NK-TR protein).  
GN NKTR.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Blood.  
RX MEDLINE=93133824; PubMed=8421688;  
RA Anderson S.K., Gallinger S., Roder J., Frey J., Young H.A., Ortaldo J.R.;  
RT "A cyclophilin-related protein involved in the function of natural killer cells."  
RL Proc. Natl. Acad. Sci. U.S.A. 90:542-546 (1993).  
RN [2]  
RP REVISIONS.  
RA Anderson S.K.;  
RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.  
CC -I- FUNCTION: Component of a putative tumor-recognition complex. Involved in the function of NK cells.  
CC -I- SUBCELLULAR LOCATION: Membrane-anchored. Attached to the membrane via its N-terminus.



CC CC -1- SIMILARITY: Contains 1 cyclophilin-like ppiase domain.

CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL; L04288; AAA35734.2; -

DR EMBL; AF184110; AAD56402.1; -

DR PIR; A47328; A47328.

DR HSSP; Q27450; IA33.

DR Genew; HGNC:7833; NKTR.

DR MIM; 161565; -

DR GO; GO:0016018; F:cyclosporin A binding; TAS.

DR InterPro; IPR002130; CSA\_PPIase.

DR Pfam; PF00160; pro\_isomerase; 1.

DR PRINTS; PR00153; CSAPPISMRASE.

DR PROSITE; PS00170; CSA\_PPIASE\_1; 1.

DR PROSITE; PS50072; CSA\_PPIASE\_2; 1.

DR Cyclosporin; Isomerase; Rotamase; Repeat; Membrane.

DR DOMAIN 1 176 PPIASE, CYCLOPHILIN-TYPE.

FT DOMAIN 219 240 ARG/LYS-RICH (BASIC).

FT DOMAIN 421 457 ARG/LYS-RICH (BASIC).

FT DOMAIN 970 1010 ARG/LYS-RICH (BASIC).

FT DOMAIN 194 244 ARG/SER-RICH.

FT DOMAIN 466 574 ARG/SER-RICH.

FT DOMAIN 664 814 ARG/SER-RICH.

FT DOMAIN 1311 1348 ARG-SER TANDM REPEAT-RICH.

SQ SEQUENCE 1462 AA; 165676 MW; D98A1147763EF527 CRC64;

## Alignment Scores:

Pred. No.: 0.00337 Length: 1462

Score: 174.50 Matches: 158

Percent Similarity: 33.84% Conservative: 110

Best Local Similarity: 19.95% Mismatches: 298

Query Match: 2.41% Indels: 226

DB: 1 Gaps: 32

US-09-302-812-1 (1-4070) x NKCR\_HUMAN (1-1462)

QY 113 AGCAGAGCTGCAGAGCACTAGCGCGAGGGGCGATGGTCCCGAGGCGCCAGGA 172

Db 512 ThrHisSerSerArgSerTyrArgSerHisSerHisSerGlnSerTyrSerArgGly 531

QY 173 GGGGGCGCAGTCGCTCCCTCCAGGGTTAGTGAATGAGGCTCTACGCCCGGCTGCCCG 232

Db 532 -SerSerArgSer-----ArgThrAlaSerLysSerSerHisSerArgSerAr 548

QY 233 GAGACTCAGTCTGGGGTCCAGCATGAGTGGCGGGCGCGGTGTGAGCCCTGCACCA 292

Db 548 gSerLysSerArg-----SerSerLysSerGlyHisArgLysArgAlaSerLy 565

QY 293 GCGACCCCGCTGGGACGCGCT-----GCACTTCTCCGCC 328

Db 565 sSerProArgLysThrAlaSerGlnLeuSerGluAsnLysProValLysThrGluProle 595

QY 329 GCGCGGCTCGAGCGCGGAGC-----TTCCCGCGGCGAGGAG 367

Db 585 uArgAlaThrMetAlaGlnAsnGluAsnValValGlnProValValAlaGluAsnIl 605

QY 368 GCGGTCCTCGATCCAGACGCTCGGTGCGATTCAGGGTCCCGCGCTCCGTCAGG 427

Db 605 eProValIleProLeuSerAspSerProProSerArgTyrLysProGlyGlnLysPr 625

QY 428 CTGC-----GCCCTGGCGGGGCGAGCACCGAGGCGCGCCACCTCTCTGTTT 481

Db 625 ofTrpLysProSerTyrGluArgIleGlnGluMetLysAlaLysThrHisLeuLeuPr 645

QY 482 CAACACAGAACACTATAACCAAGTTGGATGGACACTAAAGGAATCAAGACAGTTGAATCAGA 541

Db 645 oiledGlnSerThrTyr---SerLeuAlaAsnIleLys-----GluThrGlySerSerSe 662

QY 542 AGTTTGCATAGTAAGAAACAACAATACAAGAGAGATCCATGATCAGTCTGTGTA-- 599

Db 662 rSerTyrHisLysArgGluLysAsnSerGluSerAspGlnSerThrTyrSerLysTyrSe 682

QY 600 -----CAAAAGATAAATCTTTA 616

Db 682 rAspArgSerSerGluSerSerProArgSerArgSerSerArgSerArgSerTy 702

QY 617 TCAACATAACATGGAATAATAGAAAATGTT---TCTCAGCTAGGTTTTTGATAAGTACC 673

Db 702 rSerArgSerTyrThrArgSerArgSerAlaSerHisSerHisSerArgSerSerPr 722

QY 674 AGTTGAAAAGGT-----ACACAGTATTGAAGCAGCATCAGACTCGCGCTATGTG 724

Db 722 oSerSerArgSerHisSerArgAsnLysTyrSerAspHisSerGlnCysSerArgSerSe 742

QY 725 TAAGTGG-----CAGATGAAGGGCCACACTCAGAACCGCTTTTGGAAAGTGA 772

Db 742 rSerTyrThrSerIleSerSerAspAspGlyArgAlaLysArgAlaSerSe 762

QY 773 ACTCCAGCGGTAACTCGTACAGAGCAGCTTCAATGCTAATGCTAATGCTC----- 821

Db 762 rGlyLysLysAsnSerValSerHisLysLysHisSerSerSerSerGluLysThrLeuHi 782

QY 822 -----GATCAGTCGTCCTCCCAAGAGGATGATCAGTACAGTACACAAA 859

Db 782 sSerLysTyrValLysGlyArgAspArgSerSerCysValArgLysTyrSerGluSerAr 802

QY 860 TAGT-----GAGGAGAGTAGAATAATCAGCAGTGTTTTGACACATGTAAAGCTTCGAA 913

Db 802 gSerSerLeuAspTyrSerSerAspSerGluGln-----SerSerValGlnAlaThrGl 820

QY 914 TCACAAAGCAGCATGGAAGATGAACAGAGCAGAGAGCCAGACCCAGCAAGTGTGG 973

Db 820 nSerAlaGlnGluLysGlu---LysGlnGlyGlnMetGluArgThrHisAsnLysGlnGl 839

QY 974 CAAGGCTTCCCATCTCGCAGAACCCCTGTCAGGGGTGTACAGGAGGAGGAGACAGACGTGT 1033

Db 839 uLys-----AsnArgGlyGluGluLysSerLysSerGl 850

QY 1034 GTCCGAGAGCCCTTGTCTCGAC-----ACTGGCTCTGAGGATGTTGGTACTGGACTGAA 1087

Db 850 uArgGluCysProHisSerLysLysArgThrLeuLysGluAsnLeuSerAspHisLeuAr 870

QY 1088 AAATGCCAACAGATGAATAGACAGAAAGTAGTCTAGGAATTCCTCCATTCAGAA 1147

Db 870 gAsnGlySerLysProLysArgLys-----AsnTyrAlaGlySerLysTr 885

QY 1148 AGAAAGTGAACCTGAGTCAACCAATGGATGTA--GATAATTCMAAAATAGTTGTCCAGGA 1204

Db 885 pAspSerGluSerAsnSerGluArgAspValThrLysAsnSerLysAsnAspSerHisPr 905

QY 1205 TTCAGAGCAGATGAAGACAAAGTCCAGGTTTGTATGAACAGAGAGATAGCAGTCTGC 1264

Db 905 oSerSerAspLysGluGluGlyGluAlaThrSerAspSerGluSerGluValSerGlul 925

QY 1265 TCAACAGCAATAAACCTTCAAGGTTCCACCAAGAGAGAGCTGACACTGAGTTGAGGAA 1324

Db 925 eHisIleLysValLysProThr-----Th 933

QY 1325 GCGGTCTCTGCTAGGGAGGTGAGATTCGATTACATTTCCAATTTGAAGAGGAGAGAG 1384

Db 933 rLysSerSer-----Th 937

QY 1385 TCGAGCTGAATGAATGATGTAAT-----GCCAACGACCTGGAGATAC 1429

Db 937 rAsnThrSerLeuProAspAspAsnGlyAlaTrpLysSerSerLysGlnArgThrSerTh 957

QY 1430 TTCTACCTCGAATGTAGTGTGAGTGCAGAAATCTTAAG----- 1463

Db 957 rSerAspSerGluGlySerCysSerAsnSerGluAsnArgGlyLysProGlnLysHi 977

```

QY 1464 -CAACATGGAGAAAGATTCTAAATCAGATCATTTTCATGAGTGCCTCAAGCAGA 1522
Db   :::::||||| :::::||||| :::::||||| :::::|||||
QY 977 sLYsHISgLYsSerLYsGLuAsnLeuLYsArgLYsGluHis-----ThrLYsLYsValLY 994
Db   :::::||||| :::::||||| :::::||||| :::::|||||
QY 1523 GGAACAAAGAAAGAAACAAATGTCGAATGAACATCAAGAACAGAAAGAGATCCCTAA 1582
Db   :::::||||| :::::||||| :::::||||| :::::|||||
QY 994 sGLuLYsLeuLYsGLyLYsLYsAspLYsLYsHISLYsAlaProLYsArgLYsGlnAlaph 1014
Db   :::::||||| :::::||||| :::::||||| :::::|||||
QY 1583 ATACATTTCCACCTCACCTTTCTCCAGATAAGAAATGGCTTGGAACTCTTATGAGGAGAT 1642
Db   :::::||||| :::::||||| :::::||||| :::::|||||
QY 1014 eHISpGlnPro----- 1018
QY 1643 GAGGAGATCCCAAGGTGCTGGATCCGGCTGCGCTCCCTTGAGACCATCTGCCATCACAC 1702
Db   ----- 1018
QY 1703 AGTGACTATTTCGGGTAGATCTTTTGGCAATAGGAGAGTTTCTTAAACCTTTTCCCAACACA 1762
Db   ----- 1019
QY 1763 TTTTAAAGATTGTGGGACAAACAGCAATGTAAGATGCCCTTGTTCAGAAACAAATTTGTA 1822
Db   :::::||||| :::::||||| :::::||||| :::::|||||
QY 1025 uGLuGLuGLuGluLeuAspAspLYsGlnValThrGlnGluSerLYsGluLYsLYsVal-- 1044
Db   :::::||||| :::::||||| :::::||||| :::::|||||
QY 1823 CCCTGTGGAAGATGAGATGCTGAGCGAGCTGCGAGCGCGGTGGGAATCTATTTCAGAC 1882
Db   :::::||||| :::::||||| :::::||||| :::::|||||
QY 1045 -----SerGLuAsnAsnGlu----- 1049
QY 1883 TGCACCTTCTCAACAGGCTCACTCGGCCCCAGAACCTGAAGATGCTATTCTGAAGTACAA 1942
Db   ----- 1050
QY 1943 TGTGGCA-----TATTCTAAGAAATGGAGCTTTACAGCTTTGATGATTTCTG 1990
Db   :::::||||| :::::||||| :::::||||| :::::|||||
QY 1059 uLYsSerSerGluGluAspLeuSerGlyLYsHISAspThrValThrValSerSerAspLe 1079
Db   :::::||||| :::::||||| :::::||||| :::::|||||
QY 1991 GGATAGGCTACTAGAAAGACAGAGCTCAACACTTGTATCAGTCCATCTTGCTGTATAT 2050
Db   :::::||||| :::::||||| :::::||||| :::::|||||
QY 1079 uAspGlnPheThrLYsAspAspSerLYsLeuSerILESerProThrAlaLeuAsnThrGl 1099
Db   :::::||||| :::::||||| :::::||||| :::::|||||
QY 2051 GGTGMAATGTCACCTCTGCTGCGCAATATTGTATACCCAGCAATACCACTCTCTGAAACA 2110
Db   :::::||||| :::::||||| :::::||||| :::::|||||
QY 1099 uGLuAsnValAla---CysLeuGlnAsnIle-----Gl 1109
QY 2111 GAAGATGAATCATTCATCAATCAATGTCCAGGAACAGAGTGCCTTTAGTCTAATGC 2170
Db   :::::||||| :::::||||| :::::||||| :::::|||||
QY 1109 nHISValGluGluSerValProAsnGlyValGluAspValLeuGlnThrAspAsnMe 1129
Db   :::::||||| :::::||||| :::::||||| :::::|||||
QY 2171 TTTCTTCTGTCAGCTTTCCAGCAGCAATGCCAATGCAATCAGATGATTCAGT----- 2225
Db   :::::||||| :::::||||| :::::||||| :::::|||||
QY 1129 tGluIleCysThr---ProAspArgSerSerProAlaLYsValGluGluThrSerProLe 1148
Db   ----- 2226
QY 2226 -----TATCCAGATATTAAAC 2240
Db   :::::||||| :::::||||| :::::||||| :::::|||||
QY 1148 uGLyAsnAlaArgLeuAspThrProAspIleAsn 1159
Db   -----

```

## RESULT 4

```

ID ATRX CAEEL
AC Q9U7E0; 002061; STANDARD; PRT; 1359 AA.
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Transcriptional regulator ATRX homolog (X-linked nuclear protein-1).
GN XNP-1 OR B0041.7.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A., AND CHARACTERIZATION.
RX MEDLINE=9935296; PubMed=10433961;

```

RA Villard L., Fontes M., Ewbank J.J.;  
 RT "Characterization of xnp-1, a Caenorhabditis elegans gene similar to  
 the human XNP/ATR-X gene."; Gene 236:13-19(1999).  
 RL [2]  
 RN SEQUENCE FROM N.A.  
 RP STRAIN=Bristol N2;  
 RC Fulton R., Wohldmann P.;  
 RA Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.  
 RL CC -1- FUNCTION: Could be a global transcriptional regulator. Modifies  
 gene expression by affecting chromatin (Potential).  
 CC CC -1- SUBCELLULAR LOCATION: Nuclear.  
 CC CC -1- SIMILARITY: Belongs to the SNF2/RAD54 helicase family.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC

EMBL; AF134186; AAD55361.1; -;  
 EMBL; AF000196; AAC24256.1; -;  
 PIR; T34036; T34036.  
 WormPep; B0041.7; CE17314.  
 InterPro; IPR001410; DEAD.  
 InterPro; IPR001650; Helicase\_C.  
 InterPro; IPR00330; SNF2\_N.  
 Pfam; PF00271; helicase\_C; 1.  
 Pfam; PF00176; SNF2\_N; 1.  
 SMART; SM00487; DEXDC; 1.  
 SMART; SM00490; HELIC; 1.  
 DR PROSITE; PS00690; DEAD\_ATP\_HELICASE; FALSE\_NEG.  
 KW DNA repair; Hydrolase; Helicase; Nuclear protein; ATP-binding;  
 KW NP BINDING.  
 FT SITE 496 503 ATP (POTENTIAL).  
 FT SITE 636 639 DEAD BOX.  
 FT DOMAIN 67 70 POLY-ASP.  
 FT DOMAIN 266 272 POLY-GLU.  
 FT DOMAIN 276 281 POLY-LYS.  
 FT DOMAIN 372 375 POLY-LYS.  
 FT DOMAIN 603 608 POLY-LYS.  
 FT DOMAIN 859 862 POLY-LYS.  
 FT CONFLICT 479 479 C -> F (IN REF. 2).  
 SQ SEQUENCE 1359 AA; 156191 MW; E84342547D4F4E64 CRC64;

## Alignment Scores:

Pred. No.:	0.00354	Length:	1359
Score:	174.00	Matches:	141
Percent Similarity:	34.73%	Conservative:	124
Best Local Similarity:	18.48%	Mismatches:	284
Query Match:	2.40%	Indels:	214
DB:	1	Gaps:	32

US-09-302-812-1 (1-4070) x ATRX\_CAEEL (1-1359)

QY 516 AAAGGATCAGACAGTTCGATCAGAAAGTTTCATAGTAAGAAACACATACAGA 575  
 Db ||||| ||||| :::::||||| ||||| :::::|||||  
 81 LysSerArgLYsArgAlaLYsSerGluSer---GluSerAspGluSerGluGlu 99  
 QY 576 GAAGAATCCATGATGATCTCTGTACAAAAAGATAAATTTTATCAACATAACATGGA 635  
 Db ||||| ||||| :::::||||| :::::||||| :::::|||||  
 100 AsArgLYsLYsSerLYsSerLYsLYsValAspGlnLYsLYsGluLYsSerLYs 119  
 QY 636 TTGAAATATTTCTCAGTACGTTTGTATAGTCCACAGTTCAGAAAGGTACACATAT 695  
 Db ||||| ||||| :::::||||| :::::||||| :::::|||||  
 120 LysLYsArgThrThrSerSerSerGluAspGluAspSerAspGluGluArgGluGln--- 138  
 QY 696 TTGAGCAGCATCAGACTCGCGGTATGTGTAAGTGGCAGATGAGGCCACACTCAGAA 755  
 Db ||||| ||||| :::::||||| :::::||||| :::::|||||  
 139 ---LysSerLYsLYsSerLYsLYsThrLYsGlnThrSerSerGluSerGlu 157



DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE AP-4 protein (Proto-oncogene AF4) (FEL protein).  
GN MLLT2 OR AF4 OR FEL.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=93281633; PubMed=8506309;  
RA Nakamura T., Alder H., Gu Y., Prasad R., Canaani O., Kamada N.,  
RA Gale R.P., Lange B., Crist W.M., Nowell P.C., Croce C.M.,  
RA Canaani E.;  
RA "Genes on chromosomes 4, 9, and 19 involved in 11q23 abnormalities in  
RT acute leukemia share sequence homology and/or common motifs";  
RL Proc. Natl. Acad. Sci. U.S.A. 90:4631-4635(1993).  
[2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=93184301; PubMed=8443374;  
RA Morrissey J., Tkachuk D.C., Milatovich A., Francke U., Link M.,  
RA Cleary M.L.;  
RA "A serine/proline-rich protein is fused to HRX in t(4;11) acute  
RT leukemias";  
RL Blood 81:1124-1131(1993).  
CC -1- SUBCELLULAR LOCATION: Nuclear (Potential).  
CC -1- DISEASE: Involved in acute leukemias through a chromosomal  
CC translocation t(4;11)(q21;q23) that involves MLLT2 and MLL/HRX.  
CC The result is a rogue activator protein.  
CC -1- SIMILARITY: Belongs to the AF4 family.  
CC -1- DATABASE: NAME=Atlas Genet. Cytogenet. Oncol. Haematol.;  
CC WWW="http://www.infobiogen.fr/services/chronocancer/Genes/AF4.html".  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/  
CC or send an email to license@sib-sib.ch).  
CC  
CC EMBL: L13773; AAA58360.1; --  
CC EMBL: L25050; AAA36642.1; --  
CC PIR: A58198; A58198.  
CC PIR: I39410; I39410.  
CC Genes; HGNC:7135; MLLT2.  
CC MIM: 159557; --  
CC GO: GO:0003700; F:transcription factor activity; TAS.  
CC InterPro: IPR007797; AF-4.  
CC Pfam: PF05110; AF-4; 1.  
CC KW Nuclear protein; Chromosomal translocation; Proto-oncogene.  
FT DOMAIN 483 492 POLY-SER.  
FT DOMAIN 835 843 POLY-SER.  
FT DOMAIN 866 869 POLY-PRO.  
FT DOMAIN 871 874 POLY-SER.  
FT DOMAIN 46 46 K -> R (IN REF. 2).  
FT CONFLICT 624 624 E -> G (IN REF. 2).  
FT CONFLICT 899 905 SASSTKS -> VPAVPRV (IN REF. 2).  
FT CONFLICT 928 929 EH -> AD (IN REF. 2).  
FT CONFLICT 999 999 I -> N (IN REF. 2).  
FT CONFLICT 1096 1096 A -> AR (IN REF. 2).  
FT CONFLICT 1140 1140 N -> I (IN REF. 2).  
FT CONFLICT 1177 1210 STNCTALNSSLVDLVHTVTRQGFQOOLQELTKTP -> RQ  
FT MCAWPSTAVNWTCTHDFVSSYKN (IN REF. 2).  
SQ SEQUENCE 1210 AA; 131421 MW; F0E334DF8F2EF04 CRC64;

Alignment Scores:  
Pred. No.: 0.00395 Length: 1210  
Score: 173.00 Matches: 227  
Percent Similarity: 32.79% Conservative: 137  
Best Local Similarity: 20.45% Mismatches: 453  
Query Match: 2.39% Indels: 293  
DB: 1 Gaps: 53

US-09-302-812-1 (1-4070) x AF4\_HUMAN (1-1210)  
QY 79 CQTCTGATAGGCGCTGGTTCGGAGGCTGTAGACAGGAGCTGCAGAACAGTGCAGCG 138  
Db 163 ArgLeuGlyGlnGluGlyPheGly-----SerSerHisHisLysLysGlyAspArgArg 180  
QY 139 GCAGAGGGGGCATGGTGGCGGG-----GAGGCACCCAGGAGGAGGGGGCGAGTCGC 186  
Db 181 AlaAspGlyAspHisCysAlaSerValThrAspSerAlaProGluArgGluLeuSerPro 200  
QY 187 -----TCCCTCCCGAGGGTTAGTG-----AATGAGGCTCTA 216  
Db 201 LeuLeuSerLeuProSerProValProProLeuSerProLeuHisSerAsnGlnGlnThr 220  
QY 217 CGCCCGGGCTGGCCCGAGACTCAGTGTGGCGTCCAGCATCAGTGCAGG-----267  
Db 221 LeuProArgThrGlnGlySerSerLysValHisGlySerSerAsnAsnSerLysGlyTyr 240  
QY 268 GCCCGGCTGTGAGCCCT-----GCACCAAGC 294  
Db 241 CysProAlaLysSerProLysAspLeuAlaValLysValHisAspLysGluThrProGln 260  
QY 295 GAC-----CCCGCTGGAGCCGCTCAACTTCTCCGCGCGCGCGCTCGGACG 342  
Db 261 AspSerLeuValAlaProAlaGlnProProSerGlnThrPhe---ProProProSerLeu 279  
QY 343 CCGGAGGCTTCCCGGAGGAGGCGGCTCTCCATCCAGGAGCGC-----TCCGG 396  
Db 280 ProSerLysSerValAla-MetGlnGlnLysProThrAlaTyrValArgProMetAspG 299  
QY 397 TGCAGTTACGGTCCCGCGCTCTCTCGAGCTGCGCCCTGGCGCGCGCGCGG-----448  
Db 299 YGlnAspGlnAlaProSerGluSerProGluLeuLysProLeuProGluAspTyrArgG 319  
QY 449 ----ACAGCAGCAGGAGCGCCACCTCTCTTTTCAACAGAGAGACATTAACAGTT 504  
Db 319 nGlnThrPheGluLys-ThrAspLeuLysValProAlaLysAlaLysLeuThrLysLeu 339  
QY 505 GGATGACACTAAGAAATCAAGACAGTGAATCAGAAAGTTTGCATAGTAAGAA---560  
Db 339 ysMetProSerGlnSerValGluGlnThrTyrSerAsnGluValHisCysValGluGlu 359  
QY 561 -----AACCAACAATA 570  
Db 359 leLeuLysGluMetThrHisSerTyrProProProLeuThrAlaHisHisThrProSerT 379  
QY 571 CAAGACAAGAATCCATGATGAGTTCTGTACAAAGATAACTTTTATCAACATAACATGG 630  
Db 379 hrAlaGluProSerLysPheProPheProThrLysAspSer-----GlnHisValSerS 397  
QY 631 AAAAATAGAAATGTTTCTCAGTAGTGTTCATAGTCACCCAGTTGAAAAAGGTACAC 690  
Db 397 erValThrGlnAsnGlnLysGln-----TyrAspThrSer-----SerLysThrH 412  
QY 691 AGTATTGAGCAGCATCAGACTCGCGCTATGTGTAG-----TGGCAGA 735  
Db 412 isSerAsnSerGlnGlnGlyThrSerSerMetLeuGluAspAspLeuGlnLeuSerAspS 432  
QY 736 ATGAAGGGCCACTCAGAACGGCTTTTGGAA-----AGTGAACCTCCA- 779  
Db 432 erGluAspSerAspSerGluGlnThrProGluLysProProSerSerSerAlaProProS 452  
QY 780 --CGGTAACTCTGTTACAGACAGTTCAGTAAATGCTAAT-----818  
Db 452 erAlaProGlnSerLeuSerLeuProValAlaSerAlaHisSerSerSerAlaGluSerG 472  
QY 819 -----GTGATCAGTCTCCCAAGGATCATCAGTGCACAAATAGTGAGGAGA 870  
Db 472 luSerThrSerAspSerAspSerSerSerAspSerGluSerGluSerSerSerAspS 492  
QY 871 GTAGAGATATCAGAGTGTTCGACATGTAAGCTTGGAAATGCAACGACGATGG 930  
Db 492 erGluGluAsnGlu-----ProLeuGluThrProAlaProGluProG 506



```

Db      1146 ValThrIleThrSerHisValLeuThrAlaPheAspLeuTrpGluGlnAlaGluAlaLeu 1165
QY      2797 ACAAGGCTTACTGTGGATTTCTTC 2820
Db      1166 ThrArgLysAsnLysGluPhePhe 1173

RESULT 6
ID      NKCR_MOUSE STANDARD; PRT; 1453 AA.
AC      F30415;
DT      01-APR-1993 (Rel. 25, Created)
DT      30-MAY-2000 (Rel. 39, Last sequence update)
DT      28-FEB-2003 (Rel. 41, Last annotation update)
DE      NK-tumor recognition protein (Natural-killer cells cyclophilin-
DE      related protein) (NK-IR protein).
GN      NKTR.
OS      Mus musculus (Mouse).
OC      Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX      NCBI_TaxID=10090;
RN      [1]
RP      MEDLINE=93113824; PubMed=8421688;
RA      Anderson S.K., Gallinger S., Roder J., Frey J., Young H.A.,
RA      Ortaldo J.R.;
RT      "A cyclophilin-related protein involved in the function of natural
RT      killer cells.";
RL      Proc. Natl. Acad. Sci. U.S.A. 90:542-546(1993).
RN      [2]
RP      REVISIONS TO C-TERMINUS.
RC      STRAIN=BALB/c; TISSUE=Blood;
RA      Anderson S.K.;
RL      Submitted (Oct-1999) to the EMBL/GenBank/DBJ databases.
CC      -!- FUNCTION: Component of a putative tumor-recognition complex.
CC      Involved in the function of NK cells.
CC      -!- SIMILARITY: Contains 1 cyclophilin-like pPIase domain.
CC
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC
DR      EMBL; L04289; AAA37500.2; ALT INIT.
DR      HSSP; Q27450; 1A33.
DR      MGD; MGI:97346; Nktr.
DR      InterPro; IPR002130; CSA_PPIase.
DR      Pfam; PF00160; pro_isomerase; 1.
DR      PRINTS; PR00153; CSA_PPIASE.
DR      PROSITE; PS00170; CSA_PPIASE_1; 1.
DR      PROSITE; PS00072; CSA_PPIASE_2; 1.
KW      Cyclosporin; Isomerase; Rotanase; Repeat; Transmembrane.
FT      DOMAIN 1 176 PPIASE, CYCLOPHILIN-TYPE.
FT      DOMAIN 222 241 ARG/LYS-RICH (BASIC).
FT      DOMAIN 422 459 ARG/LYS-RICH (BASIC).
FT      DOMAIN 964 1003 ARG/LYS-RICH (BASIC).
FT      DOMAIN 198 273 ARG/SER-RICH.
FT      DOMAIN 468 565 ARG/SER-RICH.
FT      DOMAIN 658 812 ARG/SER-RICH.
FT      DOMAIN 1303 1453 ARG-SER TANDEM REPEAT-RICH.
SQ      SEQUENCE 1453 AA; 163439 MW; DF1173FFB14B283E CRC64;

Alignment Scores:
Pred. No.: 0.00517 Length: 1453
Score: 171.50 Matches: 193
Percent Similarity: 30.91% Conservative: 117
Best Local Similarity: 19.24% Mismatches: 384
Query Match: 2.37% Indels: 309
DB: 40 Gaps: 40

```

```

US-09-302-812-1 (1-4070) x NKCR_MOUSE (1-1453)
QY      101 GGAGGCTGTTCAGACAGGAGCTTCAGAACAGTCAGCGGAGGGGGGATGTCGGCGG 160
Db      508 GlySerAlaSerThrHisSerSerArgSerArgSerArgSerArgSerArgSerArgSer 527
QY      161 AGGCACCGAGGAGGGGGC-----GCAGTCCGCTCCCTCCAGGGGTAGTGAATG--- 208
Db      528 AspSerArgGlySerSerArgSerArgAlaValSerLysSerSerArgSerLysLeuAsn 547
QY      209 -----AGGCTTACGCCCGCGCTGGCCCGGAGACTCAGTCTCGCGGTCCAG 256
Db      548 ArgSerLysSerArgSerSerArgSerGlyPro-ArgArgThrSerLysSerProly 567
QY      257 CATGAGTGGGGCCCGGC-----TGAGCCCTGCACCAAGCGACC 298
Db      567 sLysProAlaGlnLeuSerGluAsnLysProValLysThrGluProLeuArgProSerVa 587
QY      299 CGCTGGGACCGCGCTGCAACTTCTCCGCGCGCGCTCGAGCGCGGAGCTTCCCGG 358
Db      587 lProGlnAsnGlyAsnValLeuValGlnProValAla-----AlaGluAsnIlePro-- 604
QY      359 CAGGCAGAGCGGCTCTCGATTCCAAGGACGCTCGGTGCGAGTTCAGGTCCCGCGCTC 418
Db      605 -----ValIleProLeuSerAspSerProProSerArgTrpLysProG 620
QY      419 CTGCTCAGGCTGC-----GCCCTGGCGCGGCGGACAGCCAGCGGCGCCACCTC 472
Db      620 yGlnLysProTrpLysProSerTyrGluArgIleGlnGluMetLysAlaLysThrHi 640
QY      473 TCTTGTGTTTCAAAACAAGACTATACACAGTTGGATGGACACTAAAGGAATCAAG---AC 529
Db      640 sLeuLeuProValGlnSerThrTyrSer-----LeuThrAsnIleLysAlaTh 656
QY      530 AGTTGAATCAGAAAGTTTCATAGTAAAGAAACAATAACAAGAGAAATCCATGAT 589
Db      656 rValSerSerSerTyrHisLysArgGluLysProSerGluSerAspGlySerAlaty 676
QY      590 GAGTTCGTACAAAAGATAACTTTTATCAACATACATGAAATATAGAAATGTTTC 649
Db      676 rSerLysTyrSer----- 680
QY      650 TCAGCTAGGTTTGTATAAGTCACAGTTGAAAAGGTACACAGTATTGAG-- 701
Db      681 -----AspArgSerSerGlySerSerGlyArgSerGlySerLysSerSerAr 696
QY      702 ----CAGCATCAGCTGCGGCTATGTAGTGGCAGAGATGAAGGCCACACTCAAGACG 757
Db      696 gSerArgSerSerSerArgSerTyrThrArgSerArgSerArgSerLeuProThrSerAr 716
QY      758 GCTTTTGGAAAGTGAACCTCCAGCGTAACCTGGTACCAGAGCAGTTCAGTAATGCTAA 817
Db      716 gSerLeuSerArgSerProSerSerArgSerHisSerProAsnLysTyrSerAspGlySe 736
QY      818 TGTCTCATGTCGTCCCAAAGGATGATCAGTGCACAAATAGTACAGAGAGTAGA-- 875
Db      736 r-----GlnHisSerArgSerSerTyrThrSerValSerSerAspGlyArgAr 754
QY      876 -----GATAATCAGCAGTTTGTGACACATGTAAG-----CTTCGAA 913
Db      754 gAlaMetPheArgSerAsnArgLysSerValThrSerHisLysArgHisArgSerAr 774
QY      914 TGCAAGACAGACGATGAAGATGAA-----CAGGCGCAGAGAACCCAGAGC---CACCA 964
Db      774 nSerGluLysThrLeuHisSerLysTyrValArgGlyArgGluLysSerSerArgHisAr 794
QY      965 GAAGTGTGCAAGGCTTGCCATCTCTCAGAAAGCCTGTGCGGGGTGTGACGAGGAGAC 1024
Db      794 glyTyrSerGluSerArgSerSerLeuAspTyrThrSerAspSerAspGlnHisVa 814
QY      1025 AGACGTGGTGTCCGAGAGGCCCTTGTGCGACACTGGCTCTGAGGATGTTGGTACTGGACT 1084
Db      814 lGlnValTyrSerAlaProGluLysGlnGlyLysValGluAlaLeuAsnAspLy 834

```







Db 1169 LysSerGlnValGlnIle---AlaGluGluLysLeuGluGlyArg-----ValAlaPro 1185  
QY 159 GGAGCACCGAGAGGGCGCAGTCGCGCTCCAGGCTTGTAGTAAGAGGCTCTACG 218  
Db 1186 LysThrValGluGluGlyGln-----ProGlnThrLysMetGluGluLeuPro 1202  
QY 219 CCCGGCTGCGCCGAGAGTC-----ACTGCTGCGGGTCCAGCATAGTGGC 266  
Db 1203 SerGlyThrProGlnSerLeuProLysGluAspLysThrThrLysThrIleLysGlu 1222  
QY 267 GCGCCCGCTGTGAGCCCTGACCAAGCGACCC-----CGCTGG 305  
Db 1223 GlnProGln---ProProCysThrAlaLysProAspGlnGluLysGluAspLysSer 1241  
QY 306 GACGCGCTGCAACTTCTCCGCGCGCTCGAGACCGCCGAGCTTCCCGCGAGCGAG 365  
Db 1242 AspThrSerSerGlnGlnProLysProGlnGlyLeuSerAspThrGlyTyr--- 1260  
QY 366 AGCGCGTCTCGATTCCAAGACGCTCGGTGCAGTTCAGGTCGCGCGCTCGCTCA 425  
Db 1261 -----SerSerAsp 1263  
QY 426 GGCTGCGCTCGCGCGCGGAGACGACCGAGCGCCACCTCTCTTCTTTCAA 485  
Db 1264 GlyIleSer-----SerSerLeuGlyGluIleProSerLeuIle--- 1276  
QY 486 CAGACACTATAACCAAGTTGGATGGACACTAAGGAATCAAGACAGTTGAATCAGAAAGT 545  
Db 1276 ----- 1276  
QY 546 TTGCATAGTAAGAAAAACAACATCAACAGAGAGATCCATGATGATCTCTACAAAA 605  
Db 1277 -----ProThrAspGluLysAspIleLeuLysGlyLeuLys 1289  
QY 606 GATAACTTTTATCAACAT-----AACATGAAAAATTAGAAAA 644  
Db 1290 AspSerPheSerGlnGluSerProSerProSerAspLeuAlaLysLeuGluSer 1309  
QY 645 -----GTTTCTCAGCTAGTGGTTTGATTAAGTCACCGATT 677  
Db 1310 ThrValLeuSerIleLeuGluAlaGlnAlaSerThrLeuAlaAspGluLysSer--- 1327  
QY 678 GAAAGGTACACAGTATTGAAGCAGCATCAGCTCGCGCTATGTTGAAGTGGCAGAT 737  
Db 1328 GluLysLysThrGln-----ProHisGluValSerPro--- 1338  
QY 738 GAAGGCCACACTCAGAACGGCTTTTGGAA-----AGTGAACCTCCAGCGTAACCTCG 791  
Db 1339 GluGlnProLysAspGlnGluLysThrGlnSerLeuSerGluThrLeuGluIleThrIle 1358  
QY 792 GTACACAGACGATTAGTAATGCTAATGTGATCGATCGTCC-----CCAAGAGTAT 845  
Db 1359 SerGluGluGluIleLysGluSerGlnGluArgLysAspThrPheLysLysAspSer 1378  
QY 846 CACAGTGAC---ACAAATAGTAGGAGAGTAGAGATAATCAGCAGTTTGTGACATGTA 902  
Db 1379 GlnGlnAspIleProSerSerLysAspHisLysGluLysGluLysPheValAspAspIle 1398  
QY 903 AAGCTTGGC-----AATGCAACAGCAGCATGGAGATGCAACAGGGC--- 944  
Db 1399 ThrThrArgArgGluProTyrAspSerValGluGluSerSerGluSerGluAsnSerPro 1418  
QY 945 -----AGAGAAGCCAGAACCCAGCAAGTGTGGCAGCGCTTGCCTCTGCAGAA 995  
Db 1419 ValProGlnArgLysArgArgThrSerValGlySerSerSerAspGluTyrLysGln 1438  
QY 996 GCCTGTGCGGGGTGACGAGGAGGAGACAGAGTGTGTCGAGAGCCCTGTGCGGAC 1055  
Db 1439 GluAspSerGlnGlySerGlyGluGluAspPheIleArgLysGlnIleIleGluMet 1458  
QY 1056 ACTGGCTCTGAGGATGTGGTACTGCACTGAAAAATGCCAACAGATTGAATAGACAGAA 1115  
Db 1115 -----LysAla 1805

Db 1459 SerAlaAspGluAsp---AlaSerGlySerGluAspAspGluPheIleArgAsnGlnLeu 1477  
QY 1116 AGTAGCTAGAGAAATTCCTCCTCATTTGAGAAAGAAAGTGAACCTGAG----- 1163  
Db 1478 LysGluIleSerSerSerThrGluSerGlnLysLysGluGluThrLysGlyLysGlyLys 1497  
QY 1164 -----TCAACCAATGATGATGAT 1181  
Db 1498 IleThrAlaGlyLysHisArgArgLeuThrArgLysSerSerThrSerIleAspGluAsp 1517  
QY 1182 AATTCCAAAAATAGTTGTGAGGATTCAGACAGATGAAGAGACAAAGTCCAGGTTTGTAT 1241  
Db 1518 AlaGlyArgArgHisSerTyrHisAspGluAspGlu-----AlaPheAsp 1533  
QY 1242 GAACAGAGAGATAGCAGTTCGCTCAACACAGCAATAAACCTTCA----- 1286  
Db 1534 GluSerProGluLeuLysTyrArgGluThrLysSerGlnGluSerGluGluLeuValVal 1553  
QY 1287 -----AGTTTCCAACCAAGAGAGCTGACACCTGAGTTGAGGAG 1325  
Db 1554 ThrGlyGlyGlyLeuArgArgPheLysThrIleGluLeuAsnSerThrIleAlaAsp 1573  
QY 1326 CGTCTCTCTGTAAGGAGGTGAGATTCGATTACATTTCCTCAATT----- 1370  
Db 1574 LysTyrSerAlaGluSerSerGlnLysLysThrSerLeuTyrPheAspGluGluProGlu 1593  
QY 1371 -----GAAGGAGGAGAG 1382  
Db 1594 LeuGluMetGluSerLeuThrAspSerProGluAspArgSerArgGlyGluGlySerSer 1613  
QY 1383 AGTCGAGCTGGAATGAATGATGTAATGCCAACCGACCTGGAAGTACTTCTAGCTGAT 1442  
Db 1614 SerLeuHisAlaSerSerPheThrProGlyThrSerProThrSerValSerSerLeuAsp 1633  
QY 1443 GTAGAGTCAGAAATCTTAAGCAACATGGGAGAAAGGATCTTAAA----- 1487  
Db 1634 GluAspAspSerSerProSerHisLysLysGlyLysGlnArgLysAla 1653  
QY 1488 ---ATCACAGATCATTTTCATGAGCTGCCAACAGCAGAGGACAAA----- 1529  
Db 1654 ArgHisArgProHisGlyProLeuLeuProThrIleGluAspSerSerGluGluGlu 1673  
QY 1530 ---AGAAAGAACAACTGTAAATGAACATCAACAGACAGAAAGGAGATC----- 1577  
Db 1674 LeuArgGluGluGluLeuLysGluGlnGluLysGlnArgGluIleGlnGln 1693  
QY 1577 ----- 1577  
Db 1694 GlnArgLysSerSerSerLysLysSerLysAspLysAspGluLeuArgAlaGlnArg 1713  
QY 1578 -----CCTAAATATACATTCACCTCACCTTCTCCAGATAAGAAATGGCTTGA 1625  
Db 1714 ArgArgGluArgProLysThrProProSerAsnLeuSerProIleGluAsp-----Ala 1731  
QY 1626 ACTCTATTGAGGAGATGAGAGAAATGCCAAGTGTGGATCCGGCTCCCTCCCTGAGA 1685  
Db 1732 SerProThrGluGluLeuArgGln-----AlaAlaGluMetGluGluLeuHis 1747  
QY 1686 CCATCT-----GCCAATCACACATGACTATTCCGGTAGATCTTTTTCGATAGGAGAA 1739  
Db 1748 ArgSerSerCysSerGluTyrSerProSerIleGluSerAsp----- 1761  
QY 1740 GTTCTTAAACCTTTCCCAACACATTTTAAAGATTGTGGAGC---AACAAAGCATGTTAAG 1796  
Db 1762 ---ProGluGlyPheGluIleSerProGluLysIleIleGluValGlnLysValTyrLys 1780  
QY 1797 ATGCTCTGTTCAGAACAAACTTGTACCTGTGGAAGATGAGATGGTGGAGGAGCTGCA 1856  
Db 1781 LeuProThrAlaValSerLeuTyrSerProThrAspGluGlnSerIleMetGlnLysGlu 1800  
QY 1857 GGCACCGGTGGGAACCTATTGACAGCTGCTCTCAACAGGCTCACTCGGCCCGCAGAAC 1916  
Db 1801 GlySerGln-----LysAla 1805

```
QY 1917 CTGAAGATGCTATTCGAAGTCAATGTCGCATATTTCTAAGAAATGGGACTTTACAGCT 1976
Db LeuLysSerAlaGluGluMetTyrGluGluMetMethHisLysThrHisLysTyrLysAla 1825
QY 1977 TTG-----ATTGATTTCTGGGTAAAGTACTAGAACAGACAGAA----- 2015
Db PheProAlaAsnGluArgAspGluValPheGluLysGluProLeuTyrGlyGlyMet 1845
QY 2016 -----GCTCAACACTTGATTCATGCTCCATCTGCCTGAT 2048
Db LeuLeuGluAspTyrIleTyrGluSerLeuValGluAsp 1858

RESULT 8
UN89 CAEEL STANDARD; PRT; 6632 AA.
AC 001761; Q17362;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Muscle M-line assembly protein unc-89 (Uncoordinated protein 89).
GN UNC-89 OR C09D1.1.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A., FUNCTION, AND TISSUE SPECIFICITY.
RC STRAIN=Bristol N2;
RX MEDLINE=96180278; PubMed=9603916;
RA Benian G.M., Tinley T.L., Tang X., Borodovsky M.;
RT "the Caenorhabditis elegans gene unc-89, required for muscle M-line
RT assembly, encodes a grant modular protein composed of Ig and signal
RT transduction domains."
RL J. Cell Biol. 132:835-848 (1996).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Du Z., Le T.T., Wilson R.;
RL Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.
RN [3]
RP REVISIONS.
RA Waterston R.;
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: Structural component of the muscle M-line. Myofilament
CC lattice assembly begins with positional cues laid down in the
CC basement membrane and muscle cell membrane. UNC-89 responds to
CC these signals, localizes, and then participates in assembling an
CC M-line.
CC -!- TISSUE SPECIFICITY: Localizes to the middle of A-bands.
CC -!- SIMILARITY: Contains 1 DBL-homology (DH) domain.
CC -!- SIMILARITY: Contains 1 fibronectin type III domain.
CC -!- SIMILARITY: Contains 49 immunoglobulin-like C2-type domains.
CC -!- SIMILARITY: Contains 1 PH domain.
CC -!- SIMILARITY: Contains 5 RCSD domains.
CC -!- SIMILARITY: Contains 1 SH3 domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U33058; AAB00542.1; -
CC DB; AF003131; AAB54132.2; -
CC PDB; 1FHO; 20-DEC-00.
CC WormPep; C09D1.1; CE30426.
DR InterPro; IPR008957; FN_III-like.
DR InterPro; IPR003961; FN_III.
DR InterPro; IPR007110; Ig-like.
DR InterPro; IPR003598; Ig_c2.

InterPro; IPR003006; Ig_MHC.
DR InterPro; IPR001849; PH.
DR InterPro; IPR007850; RCSD.
DR InterPro; IPR000219; RhoGEF.
DR InterPro; IPR001452; SH3.
DR Pfam; PF00041; fn3; 1.
DR Pfam; PF00047; Ig; 47.
DR Pfam; PF00169; PH; 1.
DR Pfam; PF05177; RCSD; 5.
DR Pfam; PF00621; RhoGEF; 1.
DR Pfam; PF00018; SH3; 1.
DR SMART; SM00408; IGC2; 23.
DR SMART; SM00325; RhoGEF; 1.
DR SMART; SM00326; SH3; 1.
DR PROSITE; PS50010; DH_2; 1.
DR PROSITE; PS50835; IG_LIKE; 49.
DR PROSITE; PS50003; PH_DOMAIN; 1.
DR PROSITE; PS50002; SH3; 1.
KW Muscle protein; Immunoglobulin domain; Repeat; SH3 domain;
KW 3D-structure.
FT DOMAIN 63 127 SH3.
FT DOMAIN 152 330 DH.
FT DOMAIN 342 498 PH.
FT DOMAIN 547 633 IG-LIKE C2-TYPE 1.
FT DOMAIN 648 736 IG-LIKE C2-TYPE 2.
FT DOMAIN 748 838 IG-LIKE C2-TYPE 3.
FT DOMAIN 946 1033 IG-LIKE C2-TYPE 4.
FT DOMAIN 1044 1132 IG-LIKE C2-TYPE 5.
FT DOMAIN 1140 1227 IG-LIKE C2-TYPE 6.
FT DOMAIN 1272 1315 THR-RICH.
FT DOMAIN 1375 1475 RCSD 1.
FT DOMAIN 1479 1585 RCSD 2.
FT DOMAIN 1597 1695 RCSD 3.
FT DOMAIN 1700 1799 RCSD 4.
FT DOMAIN 1800 1860 RCSD 5.
FT DOMAIN 1982 2067 IG-LIKE C2-TYPE 7.
FT DOMAIN 2071 2163 IG-LIKE C2-TYPE 8.
FT DOMAIN 2171 2261 IG-LIKE C2-TYPE 9.
FT DOMAIN 2269 2359 IG-LIKE C2-TYPE 10.
FT DOMAIN 2367 2455 IG-LIKE C2-TYPE 11.
FT DOMAIN 2463 2564 IG-LIKE C2-TYPE 12.
FT DOMAIN 2563 2651 IG-LIKE C2-TYPE 13.
FT DOMAIN 2657 2746 IG-LIKE C2-TYPE 14.
FT DOMAIN 2754 2858 IG-LIKE C2-TYPE 15.
FT DOMAIN 2887 2980 IG-LIKE C2-TYPE 16.
FT DOMAIN 2994 3081 IG-LIKE C2-TYPE 17.
FT DOMAIN 3087 3183 IG-LIKE C2-TYPE 18.
FT DOMAIN 3189 3280 IG-LIKE C2-TYPE 19.
FT DOMAIN 3286 3376 IG-LIKE C2-TYPE 20.
FT DOMAIN 3384 3469 IG-LIKE C2-TYPE 21.
FT DOMAIN 3482 3572 IG-LIKE C2-TYPE 22.
FT DOMAIN 3580 3667 IG-LIKE C2-TYPE 23.
FT DOMAIN 3686 3777 IG-LIKE C2-TYPE 24.
FT DOMAIN 3817 3908 IG-LIKE C2-TYPE 25.
FT DOMAIN 3920 4009 IG-LIKE C2-TYPE 26.
FT DOMAIN 4018 4106 IG-LIKE C2-TYPE 27.
FT DOMAIN 4109 4201 IG-LIKE C2-TYPE 28.
FT DOMAIN 4212 4297 IG-LIKE C2-TYPE 29.
FT DOMAIN 4302 4387 IG-LIKE C2-TYPE 30.
FT DOMAIN 4400 4485 IG-LIKE C2-TYPE 31.
FT DOMAIN 4489 4580 IG-LIKE C2-TYPE 32.
FT DOMAIN 4588 4678 IG-LIKE C2-TYPE 33.
FT DOMAIN 4681 4771 IG-LIKE C2-TYPE 34.
FT DOMAIN 4873 4961 IG-LIKE C2-TYPE 35.
FT DOMAIN 4965 5057 IG-LIKE C2-TYPE 36.
FT DOMAIN 5067 5160 IG-LIKE C2-TYPE 37.
FT DOMAIN 5171 5260 IG-LIKE C2-TYPE 38.
FT DOMAIN 5277 5366 IG-LIKE C2-TYPE 39.
FT DOMAIN 5383 5472 IG-LIKE C2-TYPE 40.
FT DOMAIN 5487 5578 IG-LIKE C2-TYPE 41.
FT DOMAIN 5595 5685 IG-LIKE C2-TYPE 42.
FT DOMAIN 5701 5790 IG-LIKE C2-TYPE 43.
FT DOMAIN 5815 5904 IG-LIKE C2-TYPE 44.
```

```
FT DOMAIN 5925 6014 IG-LIKE C2-TYPE 45.
FT DOMAIN 6038 6130 IG-LIKE C2-TYPE 46.
FT DOMAIN 6150 6239 IG-LIKE C2-TYPE 47.
FT DOMAIN 6275 6368 FIBRONECTIN TYPE-III.
FT DOMAIN 6413 6502 IG-LIKE C2-TYPE 48.
FT DOMAIN 6507 6596 IG-LIKE C2-TYPE 49.
FT DISULFID 568 621 POTENTIAL.
FT DISULFID 2908 2975 POTENTIAL.
FT DISULFID 3015 3065 POTENTIAL.
FT DISULFID 3707 3759 POTENTIAL.
FT DISULFID 3826 3890 POTENTIAL.
FT DISULFID 5092 5157 POTENTIAL.
FT DISULFID 5298 5350 POTENTIAL.
FT DISULFID 5508 5560 POTENTIAL.
FT DISULFID 5616 5669 POTENTIAL.
FT DISULFID 5722 5764 POTENTIAL.
FT DISULFID 5936 5901 POTENTIAL.
FT DISULFID 5946 5998 POTENTIAL.
FT DISULFID 6036 6171 POTENTIAL.
FT DISULFID 6421 6486 POTENTIAL.
FT CONFLICT 2137 2137 A -> P (IN REF. 1).
FT CONFLICT 2245 2247 AKA -> PRP (IN REF. 1).
FT CONFLICT 2258 2258 A -> P (IN REF. 1).
FT CONFLICT 2284 2284 E -> G (IN REF. 1).
FT CONFLICT 2297 2297 M -> I (IN REF. 1).
FT CONFLICT 3531 3531 A -> G (IN REF. 1).
FT CONFLICT 3884 3888 DAGEY -> RRRI (IN REF. 1).
FT CONFLICT 3929 3929 A -> V (IN REF. 1).
FT CONFLICT 5134 5134 A -> P (IN REF. 1).
FT CONFLICT 5145 5145 T -> S (IN REF. 1).
FT CONFLICT 5185 5185 G -> A (IN REF. 1).
FT CONFLICT 5199 5199 K -> N (IN REF. 1).
FT CONFLICT 5202 5202 L -> F (IN REF. 1).
FT CONFLICT 5213 5213 F -> L (IN REF. 1).
FT CONFLICT 6178 6178 A -> G (IN REF. 1).
FT CONFLICT 6268 6268 K -> E (IN REF. 1).
SQ SEQUENCE 6632 AA; 731665 MW; 262D3EDD62960B89 CRC64;

Alignment Scores:
Pred. No.: 0.00931 Length: 6632
Score: 170.50 Matches: 198
Percent Similarity: 35.12% Conservative: 142
Best Local Similarity: 20.45% Mismatches: 418
Query Match: 2.35% Indels: 210
DB: 1 Gaps: 47

US-09-302-812-1 (1-4070) x UN89_CAEEL (1-6632)
QY 246 GCGGTCCTCCAGCATGAGTGGCGGCGCGCTGTGAG---CCCTGCACCAAGCGACCCCGC 302
Db 1394 AlaSerProThrLysLysThrGlyGluGluValLysSerProLysGluLysSerProAla 1413
QY 303 TGGAGCGCGCTGCAACTTCTCCGCGCGCGCTCGAGCCCGGAGCTTCCCGGCGAGG 362
Db 1414 SerProThrLysLysGluLysSerProAlaGluGluValLysSer---ProThrLys 1432
QY 363 CAGAGCGCGCTCGATTCCAGACGCTCCGCTGCGATTCCAGGTCGCGCGCTCCG 422
Db 1433 LysGluLys-----SerProSerProThrLysLysGluLysGluLysSerProSerSer 1449
QY 423 TCAGGTCGCGCTCGCGCGCGCGGACAGACCCGAGCGAGCGCCACCTCTCTCTTTTC 482
Db 1450 ProThrLysLysThrGlyAspGluValLysGluLysSerProProLysSerProThrLys 1469
QY 483 AAACAGAGACTAATACCACTTGGATGGACATAAA----- 518
Db 1470 LysGluLysSerProGluLysProGluAspValLysSerProValLysLysGluLysSer 1489
QY 519 -----GGATCAAGACAGTTGATCAGAA---AGTTGCATAGTAAGAAAC 563
Db 1490 ProAspAlaThrAsnIleValGluValSerSerGluThrThrIleGluLysThrGluThr 1509
QY 564 AACAAATACAAGAGAA-----GAATCCATGATGAGTCTCTGTACAAAAAGAT 608
```

```
Db 1510 ThrMetThrThrGluMetThrHisGluSerGluGluSerArgThrSerValLysLysGlu 1529
QY 609 AACTTTTATCAACAATAACATGGAAAAATTAGAAATGTTCTCAGCTAGGTTTTCATAG 668
Db 1530 LysThrProGluLysValAspGluLysProLysSerProThrLysLys-----AspLys 1547
QY 669 TCACCACTTGAAGAAGGTACACACTATTGAAGCAGCATCAGACTGCGGCTATGTGTAA 728
Db 1548 SerProGluLysSerIleThrGluGluLysSerProValLys----- 1562
QY 729 TGGCAGAAATGAAGGCCACACTCAGAACGGCTTTTGGAAAACTGAACTCCACCGGTA 788
Db 1563 --LysGluLysSerProGluLysValGluLysProAlaSerProThrLysLysGlu 1581
QY 789 CTGGTACAGACGACTTCAGTAATGCTAATGCTCAGTCGCTCCCAAGAGGATGATCAC 848
Db 1582 LysSerProGluLys-----ProAlaSerProThrLysLysSerGluAsnGluValLys 1599
QY 849 AGTCACACAAATAGTGAAGAGTAGAGATAATCAGCAGCTTTTTCACACATGTAAAGCT 908
Db 1600 SerProThrLysLysGluLysSerProGlu---LysSerValValGluGluLysSer 1618
QY 909 GCGAATCCAAAGCAGACGATGCAA---GATGAACAGGCGCAGAGAGCCAGAGCCACG 965
Db 1619 ProLysGluLysSerProGluLysAlaAspAspLysProLysSerProThrLysLysGlu 1638
QY 966 AAGTGTGGCAGGCTTCCATCTCCAGAGCCCTGTGCAGGCTGTGCAGCAGCAGAGACA 1025
Db 1639 LysSerProGluLysSerAlaThrGluAspValLysSerProThrLysLysGluLysSer 1658
QY 1026 --GACGTGGTGTCCGAGAGCCCTTGTGCGACACTGGCTCTGAGGATGTTGGTACT--- 1079
Db 1659 ProGluLysValGluGluLysProThrSerProThrLysLysGluLysSerProThrLys 1678
QY 1080 -----GGACTGAAATGCCACAGATGTAATGATAGACAGAAAGTAGTCTAGGA 1127
Db 1679 LysThrAspAspGluValLysSerProThrLysLysGluLysSerProGluThrValGlu 1698
QY 1128 AATTCTCTCT-----CCATTTGAGAAAGATGACCTGAGTCAGTCACCAATGATGATAG 1181
Db 1699 GluLysProAlaSerProThrLysLysGluLysSerProGlu----- 1712
QY 1182 AATTCCAAAAATAGTTGTGAGGATTCAGAACACATGAGAGACAGAGTCAGGTTTGTAT 1241
Db 1713 -----LysSerValValGluLysSerProLysGluLysSerProGluLysAla 1730
QY 1242 GAACAGGAAGATAGCAGTTCTGCTCAACACAGCAATAAACCTTCAAGTTTCAACCAAGA 1301
Db 1731 GluGluLysProLysSerProThrLysLysGluLysSerProGluLysSerAlaAlaGlu 1750
QY 1302 GAAGCTGACACTGAGTTGAGGAAG---CGGTCTCTGCTAAGGAGGAGTTCGATTTCGATTA 1358
Db 1751 GluValLysSerProThrLysLysGluLysSerProGluLysSerAlaGluLysLysPro 1770
QY 1359 CATTTCCAATTTGAAGCAGGAGAGAGTCGAGCTGGAATGAATGATGTAATGCCAAACGA 1418
Db 1771 LysSerProThrLysLysGluLysSerProValLysMetAlaAspAspGluValLysSer 1790
QY 1419 CCT-----GGAAGTACTTCTAGCTCTGATGATGATGATGATGATGATGATGATGAT 1466
Db 1791 ProThrLysLysGluLysSerProGluLysValGluGluLysProAlaSerProThrLys 1810
QY 1467 CATGGAGAAAGGATCTTAAATACACAGATCATTTTCATGAGAGTGCCC----- 1514
Db 1811 LysGluLysThrProGluLysSerAlaAlaGluGluLysLysSerProThrLysLysGlu 1830
QY 1515 -----AAGCAGAGGACAAAAGAAAGAACAAATGT---GAA 1547
Db 1831 LysSerProSerSerProThrLysLysThrGlyAspGluLysLysGluLysSerProGlu 1850
QY 1548 ATGAACATCAAGACAGAAAGAGATCCCTTAATACATTCACCTTCCACCTTCTCTCA 1607
```

Db	1851	LysProGluGluLysProLysSerSerProThrThrProLysLysSerProProGly---	SerPro	1867
QY	1608	GATAAGAAATGGCTTGGAACTCCTATTGAGGAGATGAGGAGAAATGCCAAGGTGTGGGATC	1667	
Db	1870	LysLysLysLysSerLysSerPro-----GluAlaGluLysProPro-----	1883	
QY	1668	CGGCTGCTCCCTTGGAGACCATCTGCCAAATCACACACTGACTATTCCGGGTAGATCTTTTG	1727	
Db	1884	--AlaProLysLeuThrArgAspLeuLysLeuGlnThrValAsnLysThrAspLeuAla	1902	
QY	1728	CGAATAGGAGACTTCCTAAACCTTCCCAACACATTTTAAAGATTGTGGGCAACAAG	1787	
Db	1903	HisPhe---GluValValValGluHisAlaThrGluCysLysTrpPheLeuAspGlyLys	1921	
QY	1788	CATGTTAAGATGCCCTTTCAGAACAAAACCTGTACCTCTGGGAAGATGAGAAATGGTGAG	1847	
Db	1922	GluIle-----ThrThrAlaGlnGlyValThrValSerLysAspGlnPheGlu	1938	
QY	1847	-----	1847	
Db	1939	PheArgCysSerIleAspThrThrMetPheGlySerGlyThrValSerValValAlaSer	1958	
QY	1848	CGACTTCAGCGCAGCCGGTGGGAACCTATTTCAGACTGCACCT---CTCAACAGCTCACT	1904	
Db	1959	AsnAlaAlaGlySer-----ValGluThrLysThrGluLeuLysValLeuGlu	1974	
QY	1905	CGGCCCCAGAACCTGAAGGATGCTATTCTGAAGTACAATGTGGCATATTCTAAGAATGG	1964	
Db	1975	ThrProLysGlu-----	1982	
QY	1965	GACTTTACA--GCTTTGATTGATTTC-----TGGGATAAGGTACTAGAAAGA	2009	
Db	1983	GluPheThrAspLysLeuArgAspMetGluValThrLysGlyAspThrValGlnMetAsp	2002	
QY	2010	GCAGAACTCAACAC-----TTGTATCAG-----	2033	
Db	2003	ValIleAlaLeuHisSerProLeuLysTrpTyrGlnAsnGlyAsnLeuLeuGluAsp	2022	
QY	2034	-----TCCATCTTGCCTGATATGGTGAAATTCGACACTCTGCTGCACAAAT	2078	
Db	2023	GlyLysAsnGlyValThrIleLysAsnGluGluAsnLysSerLeuIleIleProAsn	2042	
QY	2079	ATTGTGTACCCAGCCAAATACCACCTCTGAAACAGAGAATGAATCATTCACCAATGTCA	2138	
Db	2043	-----AlaGlnAspSerGlyLysIleThrValGluAlaSerAsnGluValGly---Ser	2059	
QY	2139	CAGAACAGATTGCCACTCTTTAGCTAATGCTTCTTCTCTGCACG-----	2183	
Db	2060	SerGluSerSerAlaGlnLeuThrValAsnProProSerThrThrProIleValValAsp	2079	
QY	2184	TTTCCACGA-----CGCAATGCCAAGATGAATCAGAGTATTCC	2222	
Db	2080	GlyProLysSerValThrIleLysGluThrGluThrAlaGluPheLysAlaThrIleSer	2099	
QY	2223	AGTTATCCA-----GATATTAACTTCAATCGGTGGTTTGAAGACGT	2264	
Db	2100	GlyPheProAlaProThrValLysTrpThrIleAsn---GluLysIleValGlu-----	2116	
QY	2265	TCATCAGGAAACCCAGAGAAGCTTAAACCGCTCTCTGCTACTTTAGAGAGTCACAGAG	2324	
Db	2117	GluSerArgThrIleThrThrIleLysThrGluAspValThrLeuLysIleSerAsn	2136	
QY	2325	AAAAA-----CCCCTGGGTGGTGACATTCCACAGACAGAGTCTT-----GAAGAT	2372	
Db	2137	AlaLysIleGluGlnThrThrGlyThrValLysValThrAlaGlnAsnSerAlaGlyGlnAsp	2156	
QY	2373	TTTCCAGAGTGGGAA-----AGATGTGAAAAAATCTCTGCATCGCA	2411	
Db	2157	SerLysGlnAlaAspLeuLysValGluProAsnValLysAlaProLysPheLysSerGln	2176	
QY	2412	CTG-----CATGTCACCTTACGAGGT-----ACCATAGAAGGA	2444	
Db	2177	LeuThrAspLysValAlaAsnGluGluProLeuArgTrpAsnLeuGluLeuAspGly	2196	

QY	2445	AACGGCAGGCGCATCTGCATACAGGTGGATTTCGCAAAACCCTTTCGTTGGAGGTGGTTAACC	2504
Df	2197	ProSerProGly	2200
QY	2505	AGTCGAGGACTTGTCGAAGAAGTAATCCGCTTTTAACTAACCTGACCTGAGTTGATGTTTCA	2564
Df	2201	-----ThrGluValserTrirpleuueuAsnGlyGlnProLeuThrLys	2214
QY	2565	CGGCTCTTCACTGAGGTGCTGGATCACAAATGAA---TGTCTTATCAFCACAGGTACTGAG	2621
Df	2215	SerAspThrValGlnValValAspHisGlyAspGlyThrTyrrHisValThrIleAlaGlu	2234
QY	2622	CAGTACAGTGAATACACAGSC-----TATGCCGAACATACCTCGTGGGCCGGAGCCAT	2675
Df	2235	AlalysProGluMetSerGlyThrLeuThrAlalysAlalysAenAlaIla-GlyGluCy	2254
QY	2676	GAACACAGGAGCGAAAGGAGGACGACTGCGCAGAGGGCGCAGCTGAGATCGTCGCCATCGAC	2735
Df	2254	sGluThrSerAlalysValThrValAenGlyGlyAsnLysLysProGluPheValGlnAl	2274
QY	2736	GCCCTCCACTTCAGACGCTACC	2757
Df	2274	aProGlnAsnHisGluThrThr	2281
<b>RESULT 9</b>			
ID	IF38 HUMAN	STANDARD;	PRT; 913 AA.
AC	Q99613; O00215;		
DT	30-MAY-2000 (Rel. 39, Created)		
DT	30-MAY-2000 (Rel. 39, Last sequence update)		
DE	10-OCT-2003 (Rel. 42, Last annotation update)		
DE	Eukaryotic translation initiation factor 3 subunit 8 (eIF3 p110) (eIF3c).		
GN	EIF3S8.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.		
OX	NCBI_TaxID=9606;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=97150873; PubMed=8993409;		
RA	Asano K., Kinzy T.G., Merrick W.C., Hershey J.W.B.;		
RT	"Conservation and diversity of eukaryotic translation initiation		
RT	factor eIF3.",		
RL	J. Biol. Chem. 272:1101-1109 (1997).		
RN	[2]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=99425270; PubMed=10493829;		
RA	Lofthus B.J., Kim U.-J., Sneddon V.P., Kalush P., Brandon R.,		
RA	Fuhrmann J., Mason T., Crosby M.L., Barnstead M., Cronin L.,		
RA	Deslattes Mays A., Cao Y., Xu R.X., Kang H.-L., Mitchell S.,		
RA	Richler E.E., Harris P.C., Venter J.C., Adams M.D.;		
RT	"Genome duplications and other features in 12 Mb of DNA sequence from		
RT	human chromosome 16p and 16q.";		
RL	Genomics 60:295-308(1999).		
RN	[3]		
RP	SEQUENCE FROM N.A.		
RA	Kawakami T., Noguchi S., Itoh T., Shigeta K., Senba T., Matsumura K.,		
RA	Nakajima Y., Mizuno T., Morinaga M., Ota T., Suzuki Y., Obayashi M.,		
RA	Nishi T., Shibahara T., Tanaka T., Nakamura Y., Isogai T., Sugano S.;		
RT	"NEDO human cDNA sequencing project."		
RL	Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.		
RN	[4]		
RP	SEQUENCE FROM N.A.		
RC	TISSUE=Placenta;		
RX	MEDLINE=22388257; PubMed=12477932;		
RA	Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,		
RA	Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,		
RA	Altshul S.F., Zeeberg B., Buettow K.H., Schaefer C.F., Bhat N.K.,		
RA	Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,		
RA	Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,		
RA	Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,		



[illegible]

```

671 ValProPheHisLeuHisIleAsnLeuLeuGluLeuCysValTyrLeu----- 687
2829 GTTCTTCACAGAAACCTGCTGTCAGTGGCTACAGGAAACTGGGGCTGTGGTGCCTTTGGG 2888
688 -----ValSerAlaMetLeuLeuGluLeuProTyrMetAlaAlaHisGlu 702
2889 GGTGATGCTAGACTAAAGCCTTAATACAGATCCTG-----GCACCTGCTGTAGCTGAG 2942
703 SerAspAlaArgArgMetIleSerLysGlnPheHisGlnLeuArgValGlyGlu 722
2943 CGAGAGCTGGTTTATTTCACCTTTGGGACTCAGAACTCATGAGA----- 2987
723 ArgGlnProLeu-----LeuGlyProProGluSerMetArgLuiHisValValAla 739
2988 -----GACATTACAGCATCATACATTCCTCCTCAGAGAGG 3023
740 AlaSerLysAlaMetLysMetGlyAspTyrLysThrCysHisSerPheIleIleAsnGlu 759
3024 AAACCTGACTGTTGGAGAAGTATATAAGCTG----- 3053
760 LysMetAsn---GlyLysValTyrAspLeuPheProGluAlaAspLysValArgThrMet 778
3054 CTCCTACGATATTACAACTGAGAACTCAGAAACTGCTCCACCCCGGACACACATCAAG 3113
779 LeuValArgLysIleGlnGlu-----SerLeuArg 789
3114 CTTTATCCATTCATATACCATGCAGTT 3140
790 ThrTyrLeuPheThrTyrSerSerVal 798

RESULT 10
PVDA_PLAKN
ID_PVDA_PLAKN STANDARD; PRT; 1073 AA.
AC P22545;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Duffy receptor, alpha form precursor (Erythrocyte binding protein).
OS Plasmodium knowlesi.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporidia; Plasmodium.
OX NCBI_TaxID=5850;
RN [1]
R1 SEQUENCE FROM N.A.
R2 MEDLINE=92357776; PubMed=1496004;
RA Adams J.H., Sim B.K., Dolan S.A., Fang X., Kaslow D.C., Miller L.H.;
RT "A family of erythrocyte binding proteins of malaria parasites.";
RL Proc. Natl. Acad. Sci. U.S.A. 89:7085-7089(1992).
RN [2]
R1 SEQUENCE OF 296-1073 FROM N.A.
R2 MEDLINE=91004213; PubMed=2170017;
RA Adams J.H., Hudson D.E., Torii M., Ward G.E., Welles T.E.,
RA Aikawa M., Miller L.H.;
RT "The Duffy receptor family of Plasmodium knowlesi is located within
RT the micronemes of invasive malaria merozoites.";
RL Cell 63:141-153(1990).
CC CC
CC -1- FUNCTION: Binds to the human erythrocytes Duffy blood group
CC determinant.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- SIMILARITY: HIGH, TO P.VIVAX DUFFY RECEPTOR.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M90466; AAA29602.1; -
DR EMBL; M68517; AAA29590.1; -
DR EMBL; M68518; AAA29591.1; -
DR PR; A35970; A35970.
DR InterPro; IPR008602; Duffy binding

```

DR Pfam: PF05424; Duffy binding; 1.  
KW Malaria; Receptor; Glycoprotein; Signal; Transmembrane;  
KW Multigene family.  
FT SIGNAL 1 21 POTENTIAL.  
FT CHAIN 22 1073 DUFFY RECEPTOR, ALPHA FORM.  
FT DOMAIN 22 1007 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 1008 1029 POTENTIAL.  
FT DOMAIN 1030 1073 CYTOPLASMIC (POTENTIAL).  
FT CARBOHYD 134 134 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 179 179 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 202 202 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 252 252 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 348 348 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 679 679 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 746 746 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 779 779 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 788 788 N-LINKED (GLCNAC. . .) (POTENTIAL).  
SQ SEQUENCE 1073 AA; 120683 MW; 3965FC9F46B71808 CRC64;

Alignment Scores:  
Pred. No.: 0.00781 Length: 1073  
Score: 168.00 Matches: 117  
Percent Similarity: 32.92% Conservative: 68  
Best Local Similarity: 20.82% Mismatches: 215  
Query Match: 2.32% Indels: 162  
DB: 1 Gaps: 22

US-09-302-812-1 (1-4070) x PVDA\_PLAKN (1-1073)

QY	9	GTGAACGAAGCCGAATCAGAACGGCTCAT-----	38
DB	506	ValGluGluAlaArgLysAsnThrGlnGluAsnValLysAsnValGlySerGlyValGlu	525
QY	39	---CCTGAGGCTGTGGTGGCGTGAAGAGGA---AGGCAGGGCTGTGATAGGC	92
DB	526	SerLysAlaAlaSerSerAsnProIleThrGluAlaValLysSerSerGlyGluGly	545
QY	93	CTGGTTCGGAGGCTGTGAGCAGGAGCTGCAGACAGCTCAGC---GGCAGAGGGGGC	149
DB	546	LysValGlnGlu-----AspSerAlaHisLysSerValAsnLysGlyGluGlyLys	562
QY	150	ATGGTCCGGGAGGACCCAGGAGGGGGCGAGTCCGTCCTCCCTCC-----AGGGTT	200
DB	563	SerSerThrAsnGluAlaAspProGlySerGlnSerGlyAlaProAlaSerArgSerVal	582
QY	201	AGTGAATGAGCTCTACGCCGGGCTGCCCGGAGACTCAGTGTGGGTCCCGAGCATG	260
DB	583	AspGlu-----LysAlaGlyValProAlaLeu	591
QY	261	AGTGGGGCCCGCTGTGAGCCCTGCACCAAGCGACCCGCTGGGCGCGCTGCAACT	320
DB	592	SerAlaGlyGlnGlyHisAsp-----LysValProProAlaGluAlaAlaThr	608
QY	321	TCTCCGCGCGGCTCGGACGCCCGGAGCTTCCCGCAGCAGAGCGCGTCTCGAT	380
DB	609	-----GluSerAlaValLeuHis	614
QY	381	TCCAGAGAGCTCCGGTCCAGTTCAGGTCCCGGCTCCGTCCTCGTCAGGCTGGCCCTGGC	440
DB	615	SerAlaAspLysThrProAsnThrValThrGluGluAsnLysGluGlyThrGlnMetAsp	634
QY	441	CGGGGGGACAGCACCGGAGGAGCGCCACCTCTCTTGTGTTTCAAACAGAGACTATACC	500
DB	635	GlyAlaAlaGlyLysAspGlyLysAlaProGly-----	645
QY	501	AGTTGGATGGACACTAAAGGAATCAAGACAGTTGAATCAGAA-----AGTTTGCATAGT	554
DB	646	-----ProThrValSerSerAspValProSerValGlyGly	657
QY	555	AAAGAAAACAACAAATACAGAGAAATCCATGATGAGTTCTGTACAAAAGATCACTTT	614
DB	658	LysAspSerGlyProSerThrSerAlaSer-----HisAlaLeuAlaGlyGluAsnGly	675

QY	615	TATCAACATAACATGGAAAAAATAGAAAATGTTTCTCAGCTAGGTTTTCATAGTCACCA	674
DB	676	GluValHisAsn-----GlyThrAspThrGluPro	685
QY	675	GTTGAAAAAGGTACACAGTATTTGAAGCAGCATCAGCTCGGCTATGTGTAAGTGGCAG	734
DB	686	LysGluAspGlyGlyLysAlaAspProGlnLysAspLysGlyValLysGlyLysGlnAsp	705
QY	735	AATGAA-----GGGCACACTCA-----	752
DB	706	ThrAspAspArgSerGlnGlySerLeuGlyProHisThrAspGluArgAlaThrLeuGly	725
QY	753	GAACGGCTTTTGAAGTGAACCTCAGCGGTA-----ACTCTGTACACAGAGCAG	803
DB	726	GluThrHisMetGluLysAspThrGluThrAlaGlySerThrLeuThrProGluGln	745
QY	804	-----TTCAGTAATGCTAATGTGCTGCTCAGTCGTCCTCCCAAGAGATGATCAC	848
DB	746	AsnValSerValAlaSerAspAsnGlyAsnValProGlySerGlyAsnLysGlnAsnGlu	765
QY	849	AGTGACACA-----AATAGTGAGGAGATGAGATATCAGCAGCTTTTGTACACAT	899
DB	766	GlyAlaThrAlaLeuSerGlyAlaGluSerLeuLysSerAsnGluSer-----ValHis	783
QY	900	GTAAGCTTGGCAATGCAAGCAGCAGATGGAAGATGAACAGCGGCAGAGACCAAGC	959
DB	784	LysThrIleAspAsnThrThrHisGlyLeuGluAsnLysAsnGlyGlyAsnGluLysAsp	803
QY	960	CACCAAGATGTGGCAAGCTTCCCATCTCGCAGAGACCTGTGCGAGGCTGTACAGGAG	1019
DB	804	PheGlnLys-----	806
QY	1020	GACACAGACGTGTGTCCGAGACCCCTTGTCCGACCTGCTCGGCTCTCAGGATGTTGTACT	1079
DB	807	---HisAspPheMetAsnAsnAspMetLeuAsnAspGlnAlaSerSerAsp-----	822
QY	1080	GGACTGAAAAATGCCAACAGATTGAATAGACAAGAAAGTAGTCTAGGAATTTCTCTCCA	1139
DB	822	-----	822
QY	1140	TTTGAGAAAGAAAGTGAACCTGAGTCACCAATGATGTAGATAATATTCAAAAATAGTTGT	1199
DB	823	-----HisThrSerSerAspGlnThrSerSerAspHisThrSer	835
QY	1200	CAGGATTCAGACGATGAAGAGACAAAGTCCAGTTTTCATGAACAGGAGATAGCAGT	1259
DB	836	SerAspGlnThrSerSerAspHisThrSerSerAspHisThrSerSerAspGlnThrSer	855
QY	1260	TCTGCTCAACACAGCAATAAACCTTCAGGTTCCAACCAAGAGAGCTGACACTGAGTTG	1319
DB	856	SerAspGlnThrSerSer-----AspGlnThrIleAspThrGluGly	869
QY	1320	AGGAAGCGGCTCTGCTGCTAAGGGAGGTGAGATTTCGATTACATTTTCAATTTGAAGAGGA	1379
DB	870	HisHisArgAspAsnValArgAsnProGluLysSerSerGluAspMetSerLysGly	889
QY	1380	GAGAGTCGAGCTGGAATGAATGATGTGAATGCCAAACGACCTGGAAGTACTTCTAGCTG	1439
DB	890	AspPhe-----MetArgAsnSerAsnSerAsnGluLeuTySerHisAsnAsnLeu	906
QY	1440	AATGTAGGTGCAAAATTTCTAAGCAACATGGGAGAAAGGATTCTAAATCACAAGATCAT	1499
DB	907	AsnAsnArgLysLeuAsnArgAspGlnTyGluHisArgAspValLysAlaThrArgGlu	926
QY	1500	TTCATGAGAGTGCACCAAGAGCAGACAAAGAAAGAAACAAATGTGAATGAACATCAA	1559
DB	927	LysIleIleLeuMetSerGluValAsnLysCysAsnAsnArgAlaSerValLysTyCys	946
QY	1560	AGAAACA 1565	
DB	947	AsnThr 948	

RESULT 11

SGL\_BOVIN  
ID SGL\_BOVIN STANDARD; PRT; 646 AA.  
AC P23389; O02707;  
DT 01-NOV-1991 (Rel. 20, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Secretogranin I precursor (Sgl) (Chromogranin B) (Cgb) [Contains: GAWK  
peptide; Secretolytin].  
GN CHGB.  
OS Bos taurus (Bovine).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;  
OC Bovidae; Bovinae; Bos.  
OX NCBI\_TaxID=9913;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Adrenal chromaffin;  
RX MEDLINE=91223091; PubMed=2025642;  
RA Bauer J.W., Fischer-Colbrie R.;  
RT "Primary structure of bovine chromogranin B deduced from cDNA  
sequence."; Biochem. Biophys. Acta 1089:124-126 (1991).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Adrenal medulla;  
RX MEDLINE=97282588; PubMed=9136897;  
RA Yoo S.H., Kang Y.K.;  
RT "Identification of the secretory vesicle membrane binding region of  
chromogranin B."; FEBS Lett. 406:259-262 (1997).  
RN [3]  
RP SEQUENCE OF 21-646 FROM N.A.  
RC TISSUE=Adrenal medulla;  
RA Grandy D.K., Leduc R., Makam H., Flanagan T., Diliberto E.J.,  
RA Thomas G., Civelli O., Viveros O.H.;  
RL Submitted (OCT-1990) to the EMBL/GenBank/DBJ databases.  
RN [4]  
RP SEQUENCE OF 634-646.  
RC TISSUE=Adrenal chromaffin;  
RX MEDLINE=95262699; PubMed=7744058;  
RA Strub J.-M., Garcia-Sablone P., Lonnig K., Taupenot L., Hubert P.,  
RA "Processing of chromogranin B in bovine adrenal medulla.  
Identification of secretolytin, the endogenous C-terminal fragment of  
residues 614-626 with antibacterial activity."; Eur. J. Biochem. 229:356-368 (1995).  
RN [5]  
RP CHARACTERIZATION OF SECRETOLYTIN.  
RX MEDLINE=96184581; PubMed=8603705;  
RA Strub J.M., Hubert P., Nullans G., Aunis D., Metz-Boutigue M.-H.;  
RA "Antibacterial activity of secretolytin, a chromogranin B-derived  
peptide (614-626), is correlated with peptide structure.";  
FEBS Lett. 379:273-278 (1996).  
CC -!- FUNCTION: Secretogranin I is a neuroendocrine secretory granule  
protein, which may be the precursor for other biologically active  
peptides. The 16 pairs of basic AA distributed throughout its  
sequence may be used as proteolytic cleavage sites.  
CC -!- FUNCTION: Secretolytin has antibacterial activity.  
CC -!- SUBCELLULAR LOCATION: Neuroendocrine and endocrine secretory  
granules.  
CC -!- PTM: O-glycosylated (Probable).  
CC -!- SIMILARITY: Belongs to the chromogranin / secretogranin protein  
family.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; X55027; CAA38846.1; -.

DR EMBL; U88551; AAC48720.1; -.  
DR EMBL; X55489; CAA39109.1; -.  
DR PIR; S15901; S15901.  
DR InterPro; IPR001819; Chromogranin AB.  
DR InterPro; IPR001990; Granin.  
DR Pfam; PF01271; Granin; 1.  
DR PRINTS; PR00659; CHROMOGRANIN.  
DR PROSITE; PS00422; GRANINS\_1; 1.  
DR PROSITE; PS00423; GRANINS\_2; 1.  
KW Sulfation; Cleavage on pair of basic residues; Signal.  
FT SIGNAL 1 20  
FT CHAIN 21 646 SECRETOGANIN I.  
FT PEPTIDE 418 484 GAWK PEPTIDE.  
FT PEPTIDE 634 646 SECRETOLYTIN.  
FT DISULFID 36 57 BY SIMILARITY.  
FT MOD\_RES 158 158 SULFATION (POTENTIAL).  
FT MOD\_RES 315 315 SULFATION (BY SIMILARITY).  
FT CONFLICT 64 64 N -> S (IN REF. 1).  
FT CONFLICT 70 70 N -> D (IN REF. 2).  
FT CONFLICT 93 98 SEAPGL -> PRSPRAS (IN REF. 3).  
FT CONFLICT 181 181 T -> M (IN REF. 2).  
FT CONFLICT 261 261 H -> R (IN REF. 2).  
FT CONFLICT 386 386 P -> R (IN REF. 2).  
FT CONFLICT 481 481 H -> L (IN REF. 3).  
FT CONFLICT 597 597 M -> V (IN REF. 2).  
SQ SEQUENCE 646 AA; 73339 MW; 420DB1178FD9B415 CRC64;  
  
Alignment Scores:  
Pred. No.: 0.0111 Length: 646  
Score: 164.50 Matches: 166  
Percent Similarity: 31.92% Conservative: 269  
Best Local Similarity: 20.70% Mismatches: 277  
Query Match: 2.27% Indels: 41  
DB: 1 Gaps: 41  
  
US-09-302-812-1 (1-4070) x SGL\_BOVIN (1-646)  
QY 183 TCGTCCCTCCCGAGGCTGTACCGCGGGCTGAGTGAATGAGTCTACCGCGGGCTGCGCCGAGACTCAGT 242  
Db 49 SerAlaProPro-----IleThrProGluCysArgGlnValLeuLys 62  
QY 243 GCTGCGCGGTCACAGCATGAGTGGCGGCGCGGCTGTGAGCCCTGCACCAAGCAGCCCGC 302  
Db 63 LysAsnGlyLysGluLeuLysAsnGluGluLysSerGlu-----AsnGluAsnThrArg 80  
QY 303 TGGGACGCGCGCTGCACTTCTCGCGCGCGCTCGGAGCGCCGCGAGCTTCCCGCGGCGAG 362  
Db 81 PheGluValArgLeuLeuArgAspProAla-----AspThrSerGluAlaProGlyLeu 98  
QY 363 CAGAGGCGCGTCTCGATTCC-----AAGAGCGCTCCGGTGCAGTTTCAGGGTCCCGCGG 416  
Db 99 SerSerArgGlu---AspSerGlyGluGlyAspAlaGlnValProThrValAlaAspThr 117  
QY 417 TCCTGTCAGCGTCGCGCTCGCGCGCGCGGCGGAGCAGCAGCGAGCGCGCCACCTCTCTT 476  
Db 118 GluSerGlyGlyHisSerArgGluArgAlaGlyGluProGlySerGlnValAla--- 136  
QY 477 GTTTTCAACAGACAGACTATACACGTTGGTGGACACTAAAGCAATCAAGACAGATTGAA 536  
Db 137 -----LysGlu 138  
QY 537 TCAGAAAGTTTGACATAGTAAGAAACAAACAATAACAGAGAGAAATCCATGATGTTCT 596  
Db 139 AlaLysThrArgTyrSerLysSerGluGlyGlnAsnArgGluGluGluMetValLys--- 157  
QY 597 GTACAAAAGATAACTTTTATCAACATAACATGAGAAAATAGAAAATGTTCTCAGCTA 656  
Db 158 -----TyrGln---LysArgGluArgGlyGluValGlySerGluGlu 170  
QY 657 GGTTTTGATAGTCCACCGTTGAAAAGGTACACAGTATTTGAAGCAGCATCAGACTCGG 716  
Db 171 ArgLeuSerGluGlyProGlyLysAlaGlnThrAlaPheLeuAsnGln----- 186



QY	717	GCTATGTGTAAGTGCAGCAATCAAGGGCCACACTCAGAACGGCTTTTGGAAAGTGAACCT	776
Db	187	-----ArgAsnGlnThrProAlaLysysGluGluLeuValSerArg---	200
QY	777	CCAGCGGTAACTCTGGTACCAGAGCAGTTCTAGTAATGCTAATGTCGATCAGTGGTCCCA	836
Db	201	-----TyrAspThrGlnSerAlaArgGlyLeuGluLysSerHisSer	214
QY	837	AAGGATGATCACAGTGCACAAATAGTCAGGAGGTAGAGATATACAGCAGTGTTCGACA	896
Db	215	ArgGluArgSerGlnGluSerGlyGluGluThrLysSerGlnGluAsnTrp-----	232
QY	897	CATGTAAAGCTTCGNAATCCAAGCAGCAGCAGTGGAGAGATGAACAGGGCAGAGAAGACCA	956
Db	233	-----ProGlnGluLeuGlnArgHisProGluGlyGlnGluAla---	245
QY	957	AGCCACCAGAGTGTGGCAAGGCTGCCATCTCGAAGAGCTGTGCGAGGTGTCAAGCAG	1016
Db	246	-----ProGlyGluSer-----	249
QY	1017	GAGGACAGACAGTGTGTGTCGAGAGCCCTTGTTCGGACACTGGCTCTGAGGATGTTGGT	1076
Db	250	---GluGluAspAla-----SerProGluValAsp-----	258
QY	1077	ACTGGACTGAAATGCCAACAGATTGAATAGACAGAAAGTAGTCTAGGAAATTCCTCT	1136
Db	259	-----LysArgHisSerArgProArgHisHisGlyArgSerArg	272
QY	1137	CCATTTGAGAAAGAAAGTAGTAACCTGAGTCAGTCCAAATGGATGTAGATAATTCMAAAATAGT	1196
Db	273	ProAspArgSerSerGlnGluGlyAsnProProLeu-----	284
QY	1197	TGTCAGGATTCAGACAGCAGATGAGAGACAGTCCAGGTTTGTGATGAACAGGAAGATAGC	1256
Db	285	-----GluGluGluSerHisValGly-----ThrGlyAsn	294
QY	1257	AGTTCCTGCTCAAAACAGCAAAATAAACCTTCAAGGTTCCAAACAGAGAAGCTCACACTGAG	1316
Db	295	SerAspGluGluLysAlaArgHisProAlaHisPheArgAlaLeuGluGlyAlaGlu	314
QY	1317	TTGAGGAAGCGTCTCTGCTTAAGGGAGGTGAGATTGCGATTATCATTTCCAAATTGAGGA	1376
Db	315	Tyr-----GlyGluGluValArgArgHisSerAla-----	324
QY	1377	GGAGAGTGCAGCTGGATGAATGATGTGAATGCCAAACGACTCCGAAAGTACTTCTAGC	1436
Db	325	-----AlaGlnAlaProGlyAspLeuGlnGly	333
QY	1437	CTGAATGTAGAGTGCAGAAATTCCTAAGCAACATGGGAGAAAGGATTCCTAAATCAGAT	1496
Db	334	AlaArgPheGlyArgGlyArgGlyGluHis-----	344
QY	1497	CATTTTCATGAGAGTCCCAAGCAGAGGACAAAGAAAGAAACAATGTGAAATGAACAT	1556
Db	345	GlnAlaLeuArgArgProSerGluGluSer-----LeuGluGlnGluAsnLysArgHis	362
QY	1557	CAAAGAACAGAAAGGAAGATCCCTAAATACATCCACTCCACCTTTCTCCAGATAAGAAA	1616
Db	363	-----GlyLeuSerProAspLeuAsn	369
QY	1617	TGGCTTGGAACTCTTATTGAGAGATG-----AGGAGATGCCAAGGTGTGGATCCGG	1670
Db	370	MetAlaGlnGlyTyrSerGluGluSerGluGluGluArgGlyProAlaProGly-----	387
QY	1671	CTGCTCCCTTGGAA-----CCATCTGCCAATCACAGTGAAGTATTCGGGTA	1718
Db	388	---ProSerTyrArgAlaArgGlyGlyGluAlaAlaAlaTyrSerThrLeuGlyGlnThr	406
QY	1719	GATCTTTTTCGGA-----ATAGGAGAAGTTCCTAAACCTTTCCCAACACATTTTAA	1769
Db	407	AspGluLysArgPheLeuGlyGlu-----ThrHisArgValGln	420
QY	1770	---GATTTGTGGGACAAAGCATGTTAAGATGCCTTGTTCAGAACAAACTGTACCT	1826

```

421 Db GluSerGlnArgAspLysAlaAaGargArgLeuProGlyGluLeuArgAsnTyrLeuAsp 440
      |||          :|||          :|||          :|||          :|||
1827 Qy GTGGAAGATGAGATGGTCAGCGAGCTGCAGGCAGCGGTGGGAACCTCATTCAGACTGCA 1886
      :|||          :|||          :|||          :|||          :|||
441 Db TyrGlyGluGlySerGlyGluGluAlaAaArgGlyIleTyrPglN----- 455
1887 Qy CTTCTCAACAGGCTCACTCGGCCCCAG---AACTCTGAAGGATGCTATTCTGAAGTACAAT 1943
      |||||          :|||          :|||          :|||          :|||
456 Db -----ProGlnGlyAspProArgAspAla----- 463
1944 Qy GTGGCATATTCTAAGAAATGGGACTTTACAGCTTTGATTGATTTCTGGGATAAGGTACTA 2003
      |||||          :|||          :|||          :|||          :|||
464 Db -----AspGluAsnArg 467
2004 Qy GNAGACGAGAGCTCAACACTTGTTATCAGTCCATCTTGCTCGATATGGTGAATAATGCA 2063
      |||||          :|||          :|||          :|||          :|||
468 Db GluGluAlaArgLeuArgGly----- 474
2064 Qy CTCGTCTGCCAAATATTGTATACCCAGCCAATACCACTCCTCGAAACAGAGAATGATCAT 2123
      |||||          :|||          :|||          :|||          :|||
475 Db -----LysGlnTyrAlaProHis 480
2124 Qy TCCATCAATGTCACAGGAACAGATTGCCAGCTCTTTTACGTAATCTTCTTCTTC----- 2177
      |||||          :|||          :|||          :|||          :|||
481 Db HisIleThr-----GluLysArgLeuGlyGluLeuLeu---AsnProPheTyrAspPro 497
      |||||          :|||          :|||          :|||          :|||
2178 Qy -----TGCACGTTTCCACGACGC----- 2195
      |||          :|||          :|||          :|||          :|||
498 Db SerGlnTyrLysSerArgPheGluArgLysAspProMetAspAspSerPheLeuGlu 517
2196 Qy -----AATGCCAAGATGAATCAGAGTATTCAGATTATTCAGATATTAAGTCTC 2243
      |||||          :|||          :|||          :|||          :|||
518 Db GlyGluGluGluAaGngLysLeuThrLeuAsnGluLysAsnPhePheProGluTyrAsnTyr 537
      |||||          :|||          :|||          :|||          :|||
2244 Qy AAT-----CGTTGTTTGA-----GCAGTTTCATCAAGGAAA 2276
      :|||          :|||          :|||          :|||          :|||
538 Db AspTyrTrpGluLysLysProPheGluGluAspValAsnTyrGlyTyrGluLysArgAsn 557
2277 Qy CCAGAGAGAGCTTAAACGCTCTCTGCTACTTTAGAAGAGTCACAGAAAAAACCACCT 2336
      |||          :|||          :|||          :|||          :|||
558 Db ProValProLysLeuAspLysLysArgGlnTyrAspArgValAlaGluLeuAspGln--- 576
2337 Qy GGGTTGGTGACATTCACAGACAGACAGCTCTTGAAGATTTCCAGAGTGGGAAAGATGTGAA 2396
      |||||          :|||          :|||          :|||          :|||
577 Db ---LeuLeuHisTyrArgLysSerAlaGlu---PheProAspPheTyrAspSerGlu 594
2397 Qy AAATCTCGACTCGACTCGATGTCATTCACGAAGGTACCATAGAGGAAAACGGCCAGGCG 2456
      :|||          :|||          :|||          :|||          :|||
595 Db GluGlnMetSerProGlnHis---ThrAlaGluAsnGluGluLysAlaGlyGlnGly 613
2457 Qy ATGCTA 2462
      :|||          :|||          :|||          :|||          :|||
614 Db ValLeu 615

RESULT 12
PVDG_PLAKN STANDARD; PRT; 1070 AA.
ID_PVDG_PLAKN
AC P50434;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Duffy receptor, gamma form precursor (Erythrocyte binding protein).
OS Plasmodium knowlesi.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporidia; Plasmodium.
OX NCBI_TaxID=5850;
RN [1]
RP SEQUENCE FROM N.A.
EX MEDLINE=92357776; PubMed=1496004;
RA Adams J.H., Sim B.K., Dolan S.A., Fang X., Kaslow D.C., Miller L.H.;
RT "A family of erythrocyte binding proteins of malaria parasites.";
RL Proc. Natl. Acad. Sci. U.S.A. 89:7085-7089(1992).
CC -!- FUNCTION: Binds to the human erythrocytes Duffy blood group

```

CC determinant  
 CC -|- SUBCELLULAR LOCATION: Type I membrane protein.  
 CC -|- SIMILARITY: HIGH, TO P.VIVAX DUFFY RECEPTOR.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL: M90695; ABA29604.1; -  
 DR InterPro: IPR008602; Duffy binding, 1.  
 DR Pfam: PF05424; Duffy binding, 1.  
 KW Malaria; Receptor; Glycoprotein; Signal; Transmembrane;  
 KW Multigene family.  
 FT SIGNAL 1 21  
 FT CHAIN 22 1070  
 FT DOMAIN 22 1003  
 FT TRANSMEM 1004 1025  
 FT DOMAIN 1026 1070  
 FT CARBOHYD 134 134  
 FT CARBOHYD 179 179  
 FT CARBOHYD 676 676  
 FT CARBOHYD 743 743  
 FT CARBOHYD 785 785  
 FT CARBOHYD 936 936  
 FT SITE 279 281  
 SQ SEQUENCE 1070 AA; 120931 MW; 703D68911BC11B50 CRC64;

Alignment Scores:  
 Pred. No.: 0.016 Length: 1070  
 Score: 163.00 Matches: 111  
 Percent Similarity: 33.15% Conservative: 69  
 Best Local Similarity: 20.44% Mismatches: 195  
 Query Match: 2.25% Indels: 168  
 DB: 1 Gaps: 22

US-09-302-812-1 (1-4070) x PVDG\_PLAKN (1-1070)

QY 57 CCGGTGGAAGAGGA---AGGACGGCTCGATAGAGCGCTGGTTCGGAG----- 104  
 DB 531 ProfileAsnGluAlaValLysSerSerGlyGluGlyLysValGlnGluAspSerAla 550  
 QY 105 -----GCTGCAGAGCAGGAGCTGCAGACAGCTCAGCGGCGAGGG-----GGCATG 152  
 DB 551 HisArgSerValAsnGluGlyGlyLysSerSerThrAsnGluAlaAspProGlySer 570  
 QY 153 GTGCGGAGGACCGGAGGAGGGGGCGGAGTCCGTCCTCCAGGGTTAGTGAATGAGGC 212  
 DB 571 GlnProGlyGly-----ProLysArgSerValAspGluLys----- 583  
 QY 213 TCTACGCCGGGCTGGCCGGAGACTCAGTGTGCGGGTCCAGCATGATGTCGGGCC 272  
 DB 584 -----AlaGlyValProAlaLeuSerAlaGlyGln 593  
 QY 273 GCGTGTGAGCCCTGCACAGGAGCCCGCTGGAGCGCGCTGCACTTCCTCGCGGCC 332  
 DB 594 GlyHisAsp-----LysValProProAlaGluAlaAlaAlaThrGluSerAlaVal 610  
 QY 333 GCGTGGAGCCCGGAGCTTCGCC-----GGCAGGAGGAGCGCGCGTC 374  
 DB 611 ProHisSerAlaAspLysThrProLysThrAlaThrGluGluAsnLysGlnArgThrGln 630  
 QY 375 CTGATTCCAAGGAGCAGCTCGGTGCGATTCAGGGTCCCGCGCTCTCGTCAGCTCGGCC 434  
 DB 631 ValAspGly-----Val 634  
 QY 435 CTGGCGGGCGGAGCAGCAGCAGGAGCGGCCACCTCTCTGTTTCAACAGAGAGACT 494  
 DB 635 AlaGlyGlyAspGlyLysAlaProGly----- 643

QY 495 ATAACCACTTGGATGGACACTAAAGGAATCAAGACAGTGAATCAGAA-----AGTTTG 548  
 DB 644 -----ProThrValSerSerAspValProSerVal 653  
 QY 549 CATAGTAAGAAACAAACAAATACAGAGAGAAATCCATGATGATGATTTCTGTACAAAAGAT 608  
 DB 654 GlyGlyLysAspSerGlyProSerThrProAlaSerHisLeuAlaGly-----Glu 670  
 QY 609 AACTTTTATCAATCAATCAATGGAATAATGAGAAATGTTTCTCAGCTAGTGGTTTGATAAG 668  
 DB 671 AsnGlyGluValHisAsn-----GlyThrAspThr 680  
 QY 669 TCACCACTGTAAGAAAGGTACACAGTATTTGAACAGCATCAGACTGCGGTATGTGTAAG 728  
 DB 681 GluProLysGluAspGlyGluLysAlaAspProGlnLysAsnLeuGluValLysGlyLys 700  
 QY 729 TGGCAGAAATGAA-----GGGCCACACTCA----- 752  
 DB 701 GlnAspThrAspAspArgSerGlnGlySerLeuGlyProHisThrAspGluArgAlaSer 720  
 QY 753 -----GAACGGCTTTTGGAAAGTGAACCTCCACCGGTA-----ACTCTGGTACCA 797  
 DB 721 LeuGlyGluThrHisMetGluLysAspThrGluThrThrGlyGlySerThrLeuThrPro 740  
 QY 798 GAGCAG-----TTTCAGTAATGCTAATGCTGATCGATCAGTCGTCCTCCCAAGGAT 842  
 DB 741 GluGlnAsnValSerValAlaSerAspAsnGlyAsnValProGlySerGlyAsnLysGln 760  
 QY 843 GATCACAGTGCACACA-----AATAGTAGAGGAGTAGAGATAATACAGCAGTCTTTTGACA 896  
 DB 761 AsnGluGlyAlaThrAlaLeuSerGlyAlaGluSerLeuGluSerGluSerVal--- 779  
 QY 897 CATGTAAAGCTTGGAAATGCAAGACAGCATCGAAGATGACAGGCGCAGAGCCAGA 956  
 DB 780 HisLysThrIleAspAsnThrHisGlyLeuGluAsnLysAsnGlyGlyAsnGluLys 799  
 QY 957 AGCCACAGAGTGTGCAAGGCTTGCATCTGCAGAGCCTGTGCAGGGTGTGCAGCAG 1016  
 DB 800 AspPheGlnLys----- 803  
 QY 1017 GAGGAGACAGAGCTGTGTCCGAGAGCCCTTGTCCGACACTGGCTCTGAGGATTTGGT 1076  
 DB 803 ----- 803  
 QY 1077 ACTGGACTGAAAATGCCACAGATTGGAATAGACAGAAAGTAGTCTAGGAAATTTCTCT 1136  
 DB 804 HisAspPheMetAsnAsnAspMetLeuAsnAspGlnThrSerSer----- 818  
 QY 1137 CCATTTGAGAAAGAAAGTAGAACCTGAGTCACCAATGGATGTAGATAATTCAAAATAGT 1196  
 DB 819 -----AspHisThrSerSerAspGlnThrSerSerAspGlnThrSerSerAspGlnThr 836  
 QY 1197 TGTCAAGGATTCAGAGCAGATGAAGACAGATCAGCGTTTGTGATGACAGAGAAATAGC 1256  
 DB 837 SerSerAspGlnThrSerSerAspGlnThrSerSerAspGlnThrSerSerAspGlnThr 856  
 QY 1257 AGTTCTGCTCAACAGCAATAAATCAAGTTCCAAAGTTCCAAAGAGAGAGTCTGACACTGAG 1316  
 DB 857 SerSerAspGlnThr-----IleAspThrGlu 865  
 QY 1317 TTGAGGAAAGCGTCTCTGTAAGGAGGAGTTCGATTCGATTCATTTCCAAATTTGAAGA 1376  
 DB 866 GlyHisHisArgAspAsnValArgAsnProGluLysSerSerGluAspMetSerLys 885  
 QY 1377 GGAGAGAGTTCGAGCTGGAAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1436  
 DB 886 GlyAspPhe-----MetArgAsnSerAsnSerAsnGluLeuTySerHisAsnAsn 902  
 QY 1437 CTGAATGTAGAGTGCAGAAATTTCAAGCAACATGGAGAGAGAGATTTCTAAATACACAGAT 1496  
 DB 903 LeuAsnAsnArgLysLeuAsnArgAspGlnTyGluHisArgAspValLysAlaThrArg 922  
 QY 1497 CATTTTCAGAGTGTGCCAAAGCAGAGGACAAAGAAAGAAACAAATGTGAATGAACAT 1556

Thu May 27 09:55:53 2004

SQ SEQUENCE 1362 AA; 152219 MW; D52EFELCF9960907 CRC64;

Db 923 GluLysLeuMetSerGluValAsnLysCysAsnAsnArgThrSerLeuLysTyr 942  
 Qy 1557 CAAAGAAC 1565  
 Db 943 CysAsnThr 945

RESULT 13

BRD4\_HUMAN STANDARD; PRT; 1362 AA.  
 AC O6085; Q96PD3;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 15-MAR-2004 (Rel. 43, Last annotation update)  
 DE Bromodomain-containing protein 4 (HUNK1 protein).  
 GN BRD4 OR HUNK1.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=21590020; PubMed=11733348;  
 RA French C.A., Miyoshi I., Aster J.C., Kubonishi I., Kroll T.G.,  
 RA Dal Cin P., Vargas S.O., Perez-Atayde A.R., Fletcher J.A.;  
 RA "BRD4 bromodomain gene rearrangement in aggressive carcinoma with  
 RT translocation t(15;19).";  
 RL Am. J. Pathol. 159:1987-1992 (2001).  
 RL [2]  
 RP SEQUENCE OF 1-722 FROM N.A.  
 RC TISSUE=Placenta;  
 RA Weber B.;  
 RL Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.  
 CC -!- SUBCELLULAR LOCATION: Nuclear (Potential).  
 CC -!- SIMILARITY: Contains 2 bromodomains.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; AF386649; AAL26987.1; --  
 DR EMBL; Y12059; CAA72780.1; --  
 DR HSP; Q92831; I891.  
 DR Genew; HGNC:13575; BRD4.  
 DR InterPro; IPR001487; Bromodomain.  
 DR Pfam; PF00439; bromodomain; 2.  
 DR PRINTS; PR00503; BROMODOMAIN.  
 DR SMART; SM00297; BROMO; 2.  
 DR PROSITE; PS00633; BROMODOMAIN 1; 1.  
 DR PROSITE; PS00614; BROMODOMAIN 2; 2.  
 DR Bromodomain; Repeat; Nuclear protein.  
 KW Bromodomain; Repeat; Nuclear protein.  
 FT DOMAIN 75 147  
 FT DOMAIN 368 440  
 FT DOMAIN 535 594  
 FT DOMAIN 692 717  
 FT DOMAIN 703 714  
 FT DOMAIN 738 743  
 FT DOMAIN 757 761  
 FT DOMAIN 764 770  
 FT DOMAIN 771 775  
 FT DOMAIN 776 783  
 FT DOMAIN 954 964  
 FT DOMAIN 974 986  
 FT DOMAIN 1011 1014  
 FT DOMAIN 1028 1033  
 FT DOMAIN 1283 1300  
 FT DOMAIN 1301 1308  
 FT DOMAIN 1335 1338  
 FT CONFLICT 720 721  
 EM -> GP (IN REF. 2).

Alignment Scores: 0.0172 Length: 1362  
 Pred. No.: 163.00 Matches: 119  
 Score: 163.00 Conservative: 100  
 Percent Similarity: 37.12% Mismatches: 228  
 Best Local Similarity: 20.17% Indels: 144  
 Query Match: 2.25% Gaps: 26  
 DB: 1

US-09-302-812-1 (1-4070) x BRD4\_HUMAN (1-1362)

Qy 156 CCGGGAGGACCGAGGAGGGGGCGCTCCGTCCTCCAGGGTTAGTGAATGAGGCTCT 215  
 Db 261 ProAlaProAlaProGlnProValGlnSerHisProIleAla-----Ala 277  
 Qy 216 ACGCCC-----GGGCTGGCCGGAGAGACTCAGTGTGCGGGTCCC 254  
 Db 278 ThrProGlnProValLysThrLysGlyValLysArgLysAlaAspThrThrPro 297  
 Qy 255 AGCATGAGTGGGGCGCGCTGTGAGCCCTGCACCAAGCGACCCCGC----- 302  
 Db 298 Thr-----ThrIleAspProIleHisGluProProSerLeuProGluProLysThr 316  
 Qy 303 -----TGGGACGCGCTGCAACTTCTCCGGCGCGCGCTCGAGCCCGGAGTTC 356  
 Db 317 LysLeuGlyGlnArgArgGluSerSerArgProValLysProProLysLysAspVal 336  
 Qy 357 GCGAGGAGAGCGGCTCTCGATTCACAGGAGCGCTCGGTGCGAGTTCAGGTC 416  
 Db 337 AspSerGlnGlnHisProAlaProGluLysSerSerLysValSerGluGlnLeuLys 356  
 Qy 417 TCCTGTCAGTGGCGCTGGCGCGCGGCGGACACCGAGCGGCGCCACCTCTCTT 476  
 Db 357 CysSer-----GlyIleLeuLysGluMet 364  
 Qy 477 GTTTTCAACACAGACATATACCACTTGG----- 506  
 Db 365 PheAlaLysLysHisAlaAlaTyrAlaTyrProPheTyrLysProValAspValGluAla 384  
 Qy 507 -----ATGCACACTAAAGGAATCAAG 527  
 Db 385 LeuGlyLeuHisAspTyrCysAspIleLysHisLysProMetAspMetSerThrLys 404  
 Qy 528 ACAGTTGAATCAGAAAGTTTGATAGTAAAGAAACAAATACAGAGAGAGATCCATG 587  
 Db 405 Ser-----LysLeuGluAlaArgGluTyrArgAspAlaGlnGluPheGlyAla 420  
 Qy 588 ATGAGTTCTGTACAAAGATTAATTTATCAACATAAC----- 626  
 Db 421 AspValArgLeuMetPheSerAsnCysTyrLysTyrAsnProProAspHisGluValVal 440  
 Qy 627 -----ATGGAATAATTAGAAATGTTCTCAGCTAGGTTTGTATAGTCCAGTTGAA 680  
 Db 441 AlaMetAlaArgLysLeuGlnAspValPheGluMetArgPheAlaLysMetProAspGlu 460  
 Qy 681 AAAGGTACACAGTATTGTGAAGCAGCATCAGCTGGCGGTATGTGTAGTGGCAGATGAA 740  
 Db 461 ProGluGluProValValAlaValSerSerProAlaVal----- 473  
 Qy 741 GGCCACACTCAGAACCGCTTTTGGAAAGTGAACCTCCAGCGGTAACTCTGTGTACAGAG 800  
 Db 474 -----ProProThrLysValValAlaProPro 483  
 Qy 801 CAGTTCAAGTAATGCTAATGTCGATCAGTCTCCCAAGGATGATCATCAGTGCACAAAT 860  
 Db 484 SerSerSerAspSerSerSerSerSerSerSerSerSerSerSerSerSerSer 502  
 Qy 861 AGTCAGAGAGAGTAGA-----GATTAATCAGCAGTTTTCACACAT 899  
 Db 503 SerGluGluGluArgAlaGlnArgLeuAlaGluLeuGlnGluLysAlaValHis 522  
 Qy 900 GTAAAGCTTGGAAATGCAAGACACCGATGCAAGGCGGACAGAGAGAGAGAGAGC 959



DB:	1	Gaps:	20
US-09-302-812-1 (1-4070) x DMP1_HUMAN (1-513)			
QY	114	GCAGGAGCTGCAGAGCAGTGCAGCGGAGAGGGGCGATGGTCCGGGAGGCACCGAGGAG	173
DB	88	AlaGlyGlyPheSerArgSerThrGlyGlyGlyGlyAspAspLysAspAspGluAsp	107
QY	174	GGGGCGCAGTCGTCCTCCCGAGGGTTAGTGAATGAGGCTTACCGCCGGGTGCCCGG	233
DB	108	-----AspSerGlyAspAspThrPheGlyAspAspAspSerGlyProGlyProLysAsp	125
QY	234	AGACTCAGTCGCGGGTCCAGCATGAGTCGGGCGCGCGTGGAGCCCTGCACCAAG	293
DB	126	ArgGlnGluGlyGlyAsnSerArgLeuGlySerAspGluAspSerAspThrIleGln	145
QY	294	CGACCCCGTCGGGAGCGCGCTCAACTTCTCCCGCGCGCGCTCGGACGCCCGGAGTTC	353
DB	146	-----AlaSerGluGluSerAlaProGlnGlyGlnAspSerAlaGlnAsp	160
QY	354	CCCGCAGCGCAGCGCGCTCTCGATTCCAGGAGCGTCCGGTCAGTTCAGGTCCTCG	413
DB	161	ThrThrSerGluSerArgLysLeuAspAsnGluAsp-----ArgValAsp	175
QY	414	CGTCTCGTCAGCTGCGCCCTGGCGCGCGGCGGACAGCACCGAGGACGCGCACCTCT	473
DB	176	SerLysProGluGly-----GlyAspSerThrGln	185
QY	474	CTTGTTTTCAACAGACACTATAACCACTGG-----	506
DB	186	-----GluSerGluSerGluGluHisTprValGlyGlyGlySerAspGlyGluSer	202
QY	507	-----ATGGACACTAAGGATCAAGACATCAAGACATTCGAATCAGAAAGT	545
DB	203	SerHisGlyAspGlySerGluLeuAspAspGluGlyMetGlnSerAspAspProGluSer	222
QY	546	TTGCATAGTAAGAAACACATACATCAAGAGATCCATGATGATGTTCTGTACAAAA	605
DB	223	IleArgSerGluArgGlyAsnSerArgMetAsnSerAlaGlyMetLysSerLysGluSer	242
QY	606	-----GATAATTTTATCAACATACATGGAATA-----TTAGAAAT	644
DB	243	GlyGluAsnSerGluGluAlaAsnThrGlnAspSerGlyGlySerGlnLeuLeuGluHis	262
QY	645	GTTTCACAGTGGTTTGTATAGTACACAGTTCGAAAGAGTACACAGTATTTGAAGCAG	704
DB	263	ProSerArgLysIlePheArgLysSerArgIleSerGlu-----	275
QY	705	CATCAGCTCGCGCTATGTGTAAGTGCAGAAATGAAGGGCCACACTCAGAACCGCTTTTG	764
DB	276	-----GluAspAspArgSerGlu-----Leu	282
QY	765	GAAAGTGAACCTCCAGCGGTAACTCTGGTACCAGAGCTTCAGTAAATGCTAAT-----	818
DB	283	AspAspAsnAsnThrMetGluGluValLysSerAspSerThrGluAsnSerAsnSerArg	302
QY	819	-----GTCATCAGTCGTCCTCCCAAGAGTATCATCAGTCACACAATAGTGGAGAGTAGA	875
DB	303	AspThrGlyLeuSerGlnProArgArgAspSerLysGlyAspSerGlnGluAspSerLys	322
QY	876	GATAATCAGCAGTTTTTTGACACATGTAAAGCTTGCGAATGCAAGACAGCATGGAAGAT	935
DB	323	GluAsn-----LeuSerGlnGluGlu	329
QY	936	GAAACAGGCGCAGAGAGCCAGAGCCACAGAGTGTGGCAAGGTTGCCATCTCTGCAGAA	995
DB	330	SerGlnAsnValAspGlyProSerSerGluSerSerGlnGluAlaAsnLeuSerSerGln	349
QY	996	GCCTGTGCAGGGTGTACAGGAGGAG-----	1022
DB	350	GluAsnSerSerGluSerGlnGluValValSerGluSerArgGlyAspAsnProAsp	369
QY	1023	-----ACAGACGTTGGTCCGAGAGCCCTTGTTCGGACACTGGCTCTGAGGATGTTGGT	1076

DB	370	ProThrThrSerTyrValGluAspGlnGluAspSerAspSerSerSerGluGluAspSerSer	389
QY	1077	ACTGGACTGAAATAATGCCACAGATTGAATAGACAAGAA-----AGTAGTCTAGGA	1127
DB	390	HisThrLeuSerHisSerLysSerGluSerArgGluGluGlnAlaAspSerGluSerSer	409
QY	1128	AATTCTCTCCATTTGAGAAAGTGAACCTGAGTCACCAATGGATGTAGATAATTCC	1187
DB	410	GluSerLeuAsnPheSerGluGluSer---ProGluSerProGluAspGluAsnSerSer	428
QY	1188	AAAAAT-----AGTTGTCAGGATTCAAGACAGATGAAGACAGCAAGTCCAGGTTTT	1238
DB	429	SerGlnGluGlyLeuGlnSerHisSerSerSerAlaGluSerGlnSerGluGluSerHis	448
QY	1239	GATCAACAGGAGGATAGCAGTCTCTCAACAGCAATAAAACCTTCAAGGTTCACAA	1298
DB	449	SerGluGluAspAspSerAspSerGlnAspSer-----SerArgSerLysGlu	464
QY	1299	AGAGAAGCTGACACTGAGTTGAGGAAGCGTCTCTGCTGAAGGGAGGTGAGATTCGATA	1358
DB	465	AspSerAsnSerThrGlu---SerLysSerSerSerGluGluAspGlyGlnLeuLys---	482
QY	1359	CATTTCCAAATTTGAAGGAGGAGAGTCCAGCTGGAATGAATGATGTGAATGCCAAACGA	1418
DB	483	AsnIleGluIleGlu-----	487
QY	1419	CCTGGAAGTACTTCTAGCCTGAATGTAGAGTCAGAAATTTCTAAGCAACATGGGAGAAG	1478
DB	488	-----SerArgLysLeuThrValAspAlaTyrHisAsnLysProIleGlyAspGln	504
QY	1479	GAT 1481	
DB	505	Asp 505	
RESULT 15			
MY9B HUMAN STANDARD; PRT; 2158 AA.			
ID	MY9B HUMAN	Q13459; O75314; Q9UNJ2; Q9UHN0;	
AC	Q13459; O75314; Q9UNJ2; Q9UHN0;	16-OCT-2001 (Rel. 40, Created)	
DT	16-OCT-2001 (Rel. 40, Last sequence update)	15-MAR-2004 (Rel. 43, Last annotation update)	
DE	Myosin IxB (Unconventional myosin-9b).	Myosin IxB OR MYR5.	
GN	Myosin IxB (Unconventional myosin-9b).	Myosin IxB OR MYR5.	
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
OX	NCBI_TaxID=9606;		
RN	[1]	SEQUENCE FROM N.A.	
RP	TISSUE=Liver, and Small intestine;		
RC	MEDLINE=97063843; PubMed=8907710;		
RX	Wirth J.A., Jensen K.A., Post P.L., Bement W.M., Mooseker M.S.;		
RA	"Human myosin-IxB, an unconventional myosin with a chimerin-like		
RT	rho/rac GTPase-activating protein domain in its tail.";		
RL	J. Cell Sci. 109:653-661(1996).		
RN	[2]	PARTIAL SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.	
RP	MEDLINE=20047919; PubMed=10580159;		
RX	Grewal P.K., Jones A.-M., Maconochie M., Lemmers R.J.F., Frants R.R.,		
RA	Hewitt J.E.;		
RT	"Cloning of the murine unconventional myosin gene Myo9b and		
RT	identification of alternative splicing.";		
RL	Gene 240:389-398(1999).		
RN	[3]	SEQUENCE OF 1940-2158 FROM N.A. (ISOFORM LONG).	
RP	MEDLINE=98158729; PubMed=9490638; M.S.;		
RX	Post P.L., Bokoch G.M., Mooseker M.S.;		
RA	"Human myosin-IxB is a mechanotchemically active motor and a GAP for		
RT	rho.";		
RL	J. Cell Sci. 111:941-950(1998).		
RN	[4]	SEQUENCE OF 1828-2023 FROM N.A. (ISOFORM SHORT).	

RC TISSUE=Placenta;  
RA Isoqai T., Ota T., Hayaishi K., Sugiyama T., Otsuki T., Suzuki Y.,  
RA Nishikawa T., Nagai K., Sugano S., Ishibashi T., Fujimori K.,  
RA Tanai H., Kimata M., Watanabe M., Hiraoka S., Ishii S., Kawai Y.,  
RA Saito K., Yamamoto J., Wakamatsu A., Nakamura Y., Nagahara K.,  
RA Masuno Y., Kanehori K.,  
RT "NEDO human cDNA sequencing project."  
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.  
CC -!- FUNCTION: Myosins are actin-based motor molecules with ATPase  
CC activity. Unconventional myosins serve in intracellular movements.  
CC May be involved in the remodeling of the actin cytoskeleton. Binds  
CC actin with high affinity both in the absence and presence of ATP  
CC and its mechanochanical activity is inhibited by calcium ions.  
CC Also acts as a GTPase activating protein on Rho.  
CC -!- SUBCELLULAR LOCATION: IN UNDIFFERENTIATED CELLS COLOCALIZES WITH  
CC F-ACTIN IN THE CELL PERIPHERY WHILE IN DIFFERENTIATED CELLS ITS  
CC LOCALIZATION IS CYTOPLASMIC WITH THE HIGHEST LEVELS IN THE  
CC PERINUCLEAR REGION.  
CC -!- ALTERNATIVE PRODUCTS:  
CC Event=Alternative splicing; Named isoforms=2;  
CC Name=Long;  
CC IsoId=Q13459-1; Sequence=Displayed;  
CC Name=Short;  
CC IsoId=Q13459-2; Sequence=VSP\_003361, VSP\_003362;  
CC -!- TISSUE SPECIFICITY: Expressed predominantly in peripheral blood  
CC leukocytes and at lower levels, in thymus, spleen, testis,  
CC prostate, ovary, brain, small intestine and lung.  
CC -!- SIMILARITY: Contains 4 IQ domains.  
CC -!- SIMILARITY: Contains 1 myosin-like globular head domain.  
CC -!- SIMILARITY: Contains 1 Ras-associating domain.  
CC -!- SIMILARITY: Contains 1 Rho-GAP domain.  
CC -!- SIMILARITY: Contains 1 zinc-dependent phorbol-ester and DAG  
CC binding domain.  
CC -!- CAUTION: The C-terminal sequence from AA 1917 onwards from Ref.1  
CC was probably a chimera.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>)  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch).  
CC  
CC -----  
CC EMBL; U42391; AAC50402.1; ALT SEQ.  
CC EMBL; AF143684; AAF00119.1; -.  
CC EMBL; AF020267; AAC26597.1; -.  
CC EMBL; AK002201; BAA92132.1; ALT\_INIT.  
CC HSP; P08799; IIMD.  
CC Genew; HGNC:7609; MYO9B.  
CC MIM; 602129; -.  
CC GO; GO:0015629; C:actin cytoskeleton; TAS.  
CC GO; GO:0005653; C:perinuclear space; TAS.  
CC GO; GO:0003774; F:motor activity; TAS.  
CC GO; GO:0008570; F:myosin ATPase activity; TAS.  
CC GO; GO:0005100; F:Rho GTPase activator activity; TAS.  
CC InterPro; IPR002219; DAG\_PE-bind.  
CC InterPro; IPR000048; IQ region.  
CC InterPro; IPR001609; myosin head.  
CC InterPro; IPR000159; RA domain.  
CC InterPro; IPR008936; RHO GAP.  
CC Pfam; PF00130; DAG\_PE-bind; 1.  
CC Pfam; PF00612; IQ; 4.  
CC Pfam; PF00063; myosin\_head; 2.  
CC Pfam; PF00788; RA; 1.  
CC Pfam; PF0620; RHO GAP; 1.  
CC PRINTS; PR00193; MYOSINHEAVY.  
CC ProDom; PD000355; myosin\_head; 1.  
CC SMART; SM00109; Cl; 1.  
CC SMART; SM00015; IQ; 4.  
CC SMART; SM00242; MYSC; 1.  
CC SMART; SM00314; RA; 1.

DR SMART; SM00324; RHO GAP; 1.  
DR PROSITE; PS00479; DAG PE BIND DOM 1; 1.  
DR PROSITE; PS00081; DAG PE BIND\_DOM\_2; 1.  
DR PROSITE; PS50096; IQ; 3.  
DR PROSITE; PS50200; RA; 1.  
DR PROSITE; PS50238; RHO GAP; 1.  
KW Myosin; Repeat; ATP-binding; Calmodulin-binding; Actin-binding;  
KW Cytoskeleton; Coiled coil; GTPase activation; Phorbol-ester binding;  
KW Zinc; Alternative splicing.  
FT DOMAIN 1 939 MYOSIN HEAD-LIKE.  
FT DOMAIN 940 1044 NECK OR REGULATORY DOMAIN.  
FT DOMAIN 1045 2158 TAIL.  
FT DOMAIN 1046 1071 COILED COIL (POTENTIAL).  
FT DOMAIN 1880 1901 COILED COIL (POTENTIAL).  
FT DOMAIN 1959 1989 COILED COIL (POTENTIAL).  
FT DOMAIN 15 114 RAS-ASSOCIATING.  
FT DOMAIN 844 855 ACTIN-BINDING.  
FT DOMAIN 957 977 IQ 1.  
FT DOMAIN 1001 1023 IQ 2.  
FT DOMAIN 1024 1053 IQ 3.  
FT DOMAIN 1633 1681 IQ 4.  
FT DOMAIN 1703 1888 PHORBOL-ESTER AND DAG BINDING.  
FT NP\_BIND 239 246 ATP (POTENTIAL).  
FT VARSPLIC 2022 2023 PP -> QY (in isoform Short).  
FT VARSPLIC 2024 2158 Missing (in isoform Short).  
FT CONFLICT 1937 1939 /FTID=VSP\_003361.  
FT CONFLICT 1947 1947 /FTID=VSP\_003362.  
FT CONFLICT 2040 2045 L -> P (IN REF. 4).  
FT CONFLICT 2049 2049 TVAAP -> PWEPLH (IN REF. 3).  
FT CONFLICT 2067 2067 P -> L (IN REF. 3).  
FT CONFLICT 2157 2158 NG -> MAESHS (IN REF. 3).  
SQ SEQUENCE 2158 AA; 243556 MW; 4978F1D770F56D28 CRC64;  
  
Alignment Scores:  
Pred. No.: 0.0227 Length: 2158  
Score: 162.00 Matches: 215  
Percent Similarity: 31.55% Conservative: 138  
Best Local Similarity: 19.21% Mismatches: 406  
Query Match: 2.24% Indels: 360  
DB: 1 Gaps: 53  
  
US-09-302-812-1 (1-4070) x MY9B\_HUMAN (1-2158)  
QY 99 CGGGAGGCTGTGAGAGGAGGCTGCGAGAGAGTCCAGCGGAGGAGGCGGCGGCGG 158  
Db 1059 ArgGluAlaLeuGluAlaAlaArgAlaGlyAlaGluGluGlyGlyGlnAlaAla 1078  
QY 159 GGAGGACACCGAGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 218  
Db 1079 GlyGlyGlnGln-----ValAlaGluGlnGlyProGlu 1089  
QY 219 CCGGGGCTGGCCCGGAGACTCAGTGTGCGGCTCCAGCATGAGTGGCGGCGGCGGCTGT 278  
Db 1090 Pro-----AlaGluAspGlyHisLeuAlaSerGluProGluVal 1103  
QY 279 GAGCCCTGACCAAGACCGCGCTGGGAGCGCGCTGCACTTCTCCGCGCGCGGCTCG 338  
Db 1104 GlnProSerAspArgSerPro---LeuGluHisSerProGluLysGluAlaProSer 1122  
QY 339 GACGCCCGGAGCTTCCCGGAGCAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 398  
Db 1123 ProGluLysThrLeuProGlnLysThrValAlaAlaGluSerHisGlu----- 1139  
QY 399 CAGTTCAGGCTCCCGGCTCTCTCTCAGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGG 458  
Db 1140 -----LysValProSerSerArgGlu-----LysArgGluSerArgArgGln 1153  
QY 459 GCGAGCGCCACTCTCTGTGTTTCAACAGAGACTATACACAGTTGGATGACACATAAA 518  
Db 1154 ArgGlyLeuGluHisValLysPheGlnAnshLysHisIleGlnSerCysLysGluSer 1173

```
QY 519 GGAATCAAG-----ACAGTTGAATCAGAAAGTTTGCATAGTAAAGAA 560
Db : : : : :
QY 1174 AlaleuargGluProSerArgValThrGlnGluGlnGlyValSerLeuLeuGluasp 1193
QY 561 AACAAACAATCAAGAGAAAGATCCATGATGATGTTCTGTACAAAAGATAACTTTTATCAA 620
Db : : : : :
QY 1194 LysLysGluSerArgGluaspGluThrLeuLeuValValGluThrGlu----- 1209
QY 621 CATAACATGGAAAAATTAGAAAATGTTTCTCAGCTAGGTTTGTATGATGTACCCAGTTGAA 680
Db : : : : :
QY 1210 -----AlaGluAsnThrSerGln-----LysGlnProThrGlu 1220
QY 681 AAAGGTACACAGTATTGAAACAGCATCAGATCGGCTATGCTAGTGGCAGAAATGAA 740
Db : : : : :
QY 1221 -----GlnProGlnAlaMetAlaValGlyValSerGluGlu 1233
QY 741 GGGCCACACTCAGAACCGCTTTTGGAAAGT----- 770
Db : : : : :
QY 1234 -----ThrGluLysThrLeuProSerGlySerProArgProGlyGlnLeuGluArg 1250
QY 771 -----GAACTCCAGCGGTAACTCTGTTACCCAGAG 800
Db : : : : :
QY 1251 ProThrSerLeuAlaLeuAspSerArgValSerProProAlaProGlySerAlaProGlu 1270
QY 801 CAGTTCACTAATGCTAATGTCGATCAGTCGTCCCAAGAGATGATCACAGTCACACAAAT 860
Db : : : : :
QY 1271 ThrProGluaspLysSerLysProCysGlySerProArgValGlnGluLysProaspSer 1290
QY 861 AGTGAGGAGAGTAGAGATAATCAGCAGCTTTTGGACACATGTAAGCTTGGCAATGCAAG 920
Db : : : : :
QY 1291 ProGlyGlySerThrGlnIleGlnArgTyrLeuAspAlaGluArgLeuAlaSerAlaVal 1310
QY 921 CAGACATGGAAGATGAACAGCGGACAGAGCCAGAGC----- 959
Db : : : : :
QY 1311 GluLeuTrpArgGlyLysLeuValAlaAlaSerProSerAlaMetLeuSerGln 1330
QY 960 -----CACCAGAAAGTGTGCAAGGCTTGCCATCTCGCAGAGGCC 998
Db : : : : :
QY 1331 SerLeuaspLeuSeraspArgHisargAlaThrGlyAlaAlaLeuThrProThrGlu--- 1349
QY 999 TGTGACGGGTGTACAGAGGAGAGACAGACGTGGTGTGCGAGAGCCCTTGTGCGACACT 1058
Db : : : : :
QY 1350 -----GluArgArgThrSerPheSerThrSeraspValSer----- 1361
QY 1059 GCCTCTGAGGATGTTGGTACTGGACTGAAATGCCAACAGATTGAATAGACAAAGAAAGT 1118
Db : : : : :
QY 1362 -----LysLeuLeuPro 1365
QY 1119 AGCTAGGAAATTCCTCCTCCTATTGAGAAAGAAAGTGAACCTGAGTCACCAATGGAT--- 1175
Db : : : : :
QY 1366 SerLeuAlaLysAlaGlnProAlaAlaGluThrThrAspGlyGluArgSerAlaLysLys 1385
QY 1176 -----GTAGATAAATCCAAAATAGTTCTCAGGATTCAGAGCAGATGAAAGACAGT 1229
Db : : : : :
QY 1386 ProAlaValGlnLysLysLysProGlyaspAlaSerSerLeuProaspAlaGlyLeuSer 1405
QY 1230 CCAGGTTTGTATCAACAGGAGATAGCAGTCTCTGCT-----CAA 1268
Db : : : : :
QY 1406 ProGly-----SerGlnValaspSerLysSerThrPheLysArgLeuPheLeuHisLys 1423
QY 1269 ACAGCAATAAACCTTCAAGGTTCCAAACCAAGAGAA----- 1304
Db : : : : :
QY 1424 ThrLysaspLysLysThrSerLeuGluGlyAlaGluLeuGluAsnAlaValSerGly 1443
QY 1305 -----GCTGACACTGAGTTGAGAAAGCGGTCTCTCTGTAAGGAGGTGAGATT 1352
Db : : : : :
QY 1444 HisValValLeuGluAlaThrThrMetLysLysGlyLeuGluAlaProSerGlyGlnGln 1463
QY 1353 CGATTACATTTCCAAATTTGAGGAGGAGAGTCTGAGCTGGAATGATGATGTAATGCC 1412
Db : : : : :
QY 1464 HisArgHis-----AlaAlaGlyGluLysArg-----Thr 1473
```

```
1413 AAACGACCTCGAAGTACTTCTAGCCTGAATGTAGAGTGCAGAAATTTCTAAGCAACATGGG 1472
Db : : : : :
QY 1474 LysGluProGlyGly-----LysGlyLysLysAsnArgAsn 1485
QY 1473 AGAAAGATTCTAAATCAACAGATCATTTTCATGAGATGCCAAAGCAGAGGACAAAAGA 1532
Db : : : : :
QY 1486 ValLysIleGlyLysIleThr-----ValSerGluLysTrp 1497
QY 1533 AAAGAAACAATGTGAATGAAACATCAAGAACACAGAAAGAGATCCCTAATAATCATTCCA 1592
Db : : : : :
QY 1498 ArgGlu-----SerValPheArgGlnIleThrAsnAlaAsnGluLeuLysTyrLeuasp 1515
QY 1593 CCTCACCTT-----TCTCCAGATAAGAAATGGCTTGGAACTCCTATTGAGGAGATG 1643
Db : : : : :
QY 1516 GluPheLeuLeuAsnLysIleAsnAspLeuArgSerGlnLysThrProIleGluSerLeu 1535
QY 1644 AGGAGAATGCCAGGTGTGGATCCGGCTGCCTCCCTTGACCATCTGCCAATCACACA 1703
Db : : : : :
QY 1536 -----PheIleGluAlaThr 1540
QY 1704 GTGACTATTCCGGTAGACTTTTTCGGAATAGGAGAGATTCCTCTAAA---CCTTTCCCAACA 1760
Db : : : : :
QY 1541 GluLysPheArgSerAsnIleLysThrMetTyrSerValProAsnGlyLysIleHisVal 1560
QY 1761 CATTTTAAAGATTTGTGGACACACATGTTAAGATGCCTTGTTCAGAACAAACATG 1820
Db : : : : :
QY 1561 GlyTyrLysaspLeuMetGluAsnTyrGlnIleValVal-----SerAsnLeu 1576
QY 1821 TACCCTGTGGAAGATGAGAATGGTGAGCGAGCTGCAGCGACCCGGTGGGAACTCATTCAG 1880
Db : : : : :
QY 1577 -----AlaThrGluArgGlyGlnLysAspThrAsnLeuValLeuAsnLeuPheGln 1593
QY 1881 ACTGCATCTTCAACAGAGCTCACTCGGCCCCAGAACCTCGAAGAT---GCTATTCTGAG 1937
Db : : : : :
QY 1594 ---SerLeuLeuaspGluPheThrArgGlyTyrThrLysAsnAspPheGluProValLys 1612
QY 1938 TACATGTGGCATATTCTAGAAATGGGACTTTACAGCTTTGATTGTTCTGCGATTAAG 1997
Db : : : : :
QY 1613 GlnSerLysAlaGlnLysLys-----Arg 1621
QY 1998 GTACTAGAAAGACAGAGAGCTCAACAC-----TTGTATCAGTCCATC 2039
Db : : : : :
QY 1622 LysGlnGluArgAlaValGlnGluHisAsnGlyHisValPheAlaSerTyrGlnValSer 1641
QY 2040 TTGCCTGAT-----ATCGTGAATTCACCTC 2066
Db : : : : :
QY 1642 IleProGlnSerCysGluGlnCysLeuSerTyrIleTrpLeuMetAspLysAlaLeuLeu 1661
QY 2067 TGTCTGCCAAATATTGTACCCAGCCATACACTCCTGMAACAGAGATGAATCATTC 2126
Db : : : : :
QY 1662 Cys-----SerValCys-----Lys 1666
QY 2127 ATCAAGTGCACAGAACAGATTCGCCAGTCTTTTTCAGCTAATGCTTTCTTGCACGTTT 2186
Db : : : : :
QY 1667 MetThrCysHisLysLysCysValHisLysIleGlnSerHisCysSerTyr---ThrTyr 1685
QY 2187 CCACGACGCAATGCC-----AAGATG 2207
Db : : : : :
QY 1686 GlyArgLysGlyGluProGlyAlaGluProGlyHisPheGlyValCysValaspSerLeu 1705
QY 2208 AAATCAGAGTATTCAGTTATCCAGATATTAACCTCAATCGCTTCTTTGAA----- 2258
Db : : : : :
QY 1706 ThrSeraspLysAlaSerValPro---IleValLeuGluLysLeuLeuGluHisValGlu 1724
QY 2259 -----GGACGTTTCATCAAGGAAACCA 2279
Db : : : : :
QY 1725 MetHisGlyLeuTyrThrGluGlyLeuTyrArgLysSerGlyAlaAlaAsnArgThrArg 1744
QY 2280 GAGAAGCTTAAACCGCTCTTCTGCTACTTTCAGAGAGTCACAGAGAAAAAACCCACTGG 2339
Db : : : : :
QY 1745 Glu-----LeuArgGlnAlaLeuGlnThrAspProAlaAla 1756
QY 2340 TTGGTGACATTCACAGACAGAGTCTTGAAGATTTTCCA----- 2378
```

Search completed: May 26, 2004, 16:49:18  
Job time : 185.05 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 26, 2004, 18:38:05 ; Search time 0.993349 Seconds  
(without alignments)  
1611.120 Million cell updates/sec

Title: US-09-302-812-11

Perfect score: 167

Sequence: 1 LFTEVLHDNECLIIITGTEQSYVTGYAETVR 31

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents AA:\*  
1: /cgn2.6/prodata/2/iaa/5A COMB.pap:\*  
2: /cgn2.6/prodata/2/iaa/5B COMB.pap:\*  
3: /cgn2.6/prodata/2/iaa/6A COMB.pap:\*  
4: /cgn2.6/prodata/2/iaa/6B COMB.pap:\*  
5: /cgn2.6/prodata/2/iaa/PCTUS COMB.pap:\*  
6: /cgn2.6/prodata/2/iaa/backfile1.pap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	% Match	Query Length	ID	Description
1	167	100.0	31	US-09-302-812-11	Sequence 11, Appl
2	167	100.0	31	US-09-511-477-11	Sequence 11, Appl
3	167	100.0	31	US-09-511-507-11	Sequence 11, Appl
4	167	100.0	968	US-09-302-812-6	Sequence 6, Appl
5	167	100.0	968	US-09-511-477-6	Sequence 6, Appl
6	167	100.0	968	US-09-511-507-6	Sequence 6, Appl
7	167	100.0	976	US-09-302-812-4	Sequence 4, Appl
8	167	100.0	976	US-09-511-477-4	Sequence 4, Appl
9	167	100.0	976	US-09-511-507-4	Sequence 4, Appl
10	167	100.0	977	US-09-302-812-2	Sequence 2, Appl
11	167	100.0	977	US-09-511-477-2	Sequence 2, Appl
12	167	100.0	977	US-09-511-507-2	Sequence 2, Appl
13	82	49.1	768	US-09-302-812-8	Sequence 8, Appl
14	82	49.1	768	US-09-511-477-8	Sequence 8, Appl
15	82	49.1	768	US-09-511-507-8	Sequence 8, Appl
16	53.5	32.0	697	US-09-540-236-3601	Sequence 3601, Ap
17	53	31.7	726	US-09-302-812-10	Sequence 10, Appl
18	53	31.7	726	US-09-511-477-10	Sequence 10, Appl
19	53	31.7	726	US-09-511-507-10	Sequence 10, Appl
20	49.5	29.6	123	US-09-107-532A-5299	Sequence 5299, Ap
21	49.5	29.6	434	US-09-107-532A-4075	Sequence 4075, Ap
22	49	29.3	453	US-08-374-155A-8	Sequence 8, Appl
23	49	29.3	453	US-08-785-396-8	Sequence 8, Appl
24	48.5	29.0	122	US-09-107-532A-5851	Sequence 5851, Ap
25	48.5	29.0	330	US-09-540-236-3751	Sequence 3751, Ap
26	48	28.7	372	US-08-597-236-8	Sequence 8, Appl
27	48	28.7	372	US-08-746-682A-8	Sequence 8, Appl

## ALIGNMENTS

## RESULT 1

US-09-302-812-11  
; Sequence 11, Application US/09302812B  
; Patent No. 6333148

GENERAL INFORMATION:

APPLICANT: JACOBSON, Myron K.

APPLICANT: JACOBSON, Elaine L.

APPLICANT: AM, Jean-Christophe

APPLICANT: LIN, Winston

TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOPOLYMERASE (PARG) ENZ

TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIVE

TITLE OF INVENTION: THEREWITH

FILE REFERENCE: NIAD 201

CURRENT APPLICATION NUMBER: US/09/302,812B

CURRENT FILING DATE: 1999-04-30

EARLIER APPLICATION NUMBER: 60/083,768

EARLIER FILING DATE: 1998-05-01

NUMBER OF SEQ ID NOS: 38

SEQ ID NO 11

LENGTH: 31

TYPE: PRT

ORGANISM: Bos taurus

FEATURE:

US-09-302-812-11

Query Match

Best Local Similarity 100.0%; Score 167; DB 4; Length 31;

Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LFTEVLHDNECLIIITGTEQSYVTGYAETVR 31

DB 1 LFTEVLHDNECLIIITGTEQSYVTGYAETVR 31

## RESULT 2

US-09-511-477-11

; Sequence 11, Application US/09511477

; Patent No. 6337202

GENERAL INFORMATION:

APPLICANT: JACOBSON, Myron K.

APPLICANT: JACOBSON, Elaine L.

APPLICANT: AME, Jean-Christophe

APPLICANT: LIN, Winston

TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOPOLYMERASE (PARG) ENZ

TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIVE

TITLE OF INVENTION: THEREWITH

FILE REFERENCE: NIAD 201

CURRENT APPLICATION NUMBER: US/09/511,477

CURRENT FILING DATE: 2000-02-23

PRIOR APPLICATION NUMBER: 09/302,812

APPLICANT: LIN, Winston  
 TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE)  
 TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIVE  
 TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIVE  
 TITLE OF INVENTION: THEREWITH  
 FILE REFERENCE: NIAID 201  
 CURRENT APPLICATION NUMBER: US/09/511,507  
 CURRENT FILING DATE: 2000-02-23  
 PRIOR APPLICATION NUMBER: 09/302,812  
 PRIOR FILING DATE: 1999-04-30  
 NUMBER OF SEQ ID NOS: 38

```

: SEQ ID NO 6
:
:   LENGTH: 968
:   TYPE: prt
:   ORGANISM: Mus musculus
:   FEATURE:
:
: IS-09-511-507-6

```

Query Match 100.0%; Score 167; DB 4; Length 968;

Best Local Similarity 100.0%; Pred. No. 9.5e-17; Indels 0; Gaps 0;  
Matches 31; Conservative 0; Mismatches 0;

QY 1 LFTVLDHNECLIIITGTEQYSEYTGVAETVR 31  
|||||  
DB 762 LFTVLDHNECLIIITGTEQYSEYTGVAETVR 792  
|||||

RESULT 7  
US-09-302-812-4  
; Sequence 4, Application US/09302812B  
; Patent No. 6333148  
; GENERAL INFORMATION:  
; APPLICANT: JACOBSON, Myron K.  
; APPLICANT: JACOBSON, Elaine L.  
; APPLICANT: AM, Jean-Christophe  
; APPLICANT: LIN, Winston  
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE (PARG) ENZ  
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIVE  
; TITLE OF INVENTION: THEREWITH  
; FILE REFERENCE: NIAD 201  
; CURRENT APPLICATION NUMBER: US/09/302,812B  
; CURRENT FILING DATE: 1999-04-30  
; EARLIER APPLICATION NUMBER: 60/083,768  
; EARLIER FILING DATE: 1998-05-01  
; NUMBER OF SEQ ID NOS: 38  
; SEQ ID NO 4  
; LENGTH: 976  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
US-09-302-812-4

Query Match 100.0%; Score 167; DB 4; Length 976;  
Best Local Similarity 100.0%; Pred. No. 9.6e-17;  
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LFTVLDHNECLIIITGTEQYSEYTGVAETVR 31  
|||||  
DB 770 LFTVLDHNECLIIITGTEQYSEYTGVAETVR 800  
|||||

RESULT 8  
US-09-511-477-4  
; Sequence 4, Application US/09511477  
; Patent No. 6337202  
; GENERAL INFORMATION:  
; APPLICANT: JACOBSON, Myron K.  
; APPLICANT: JACOBSON, Elaine L.  
; APPLICANT: AM, Jean-Christophe  
; APPLICANT: LIN, Winston  
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE (PARG) ENZ  
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIVE  
; TITLE OF INVENTION: THEREWITH  
; FILE REFERENCE: NIAD 201  
; CURRENT APPLICATION NUMBER: US/09/511,477  
; CURRENT FILING DATE: 2000-02-23  
; PRIOR APPLICATION NUMBER: 09/302,812  
; PRIOR FILING DATE: 1999-04-30  
; NUMBER OF SEQ ID NOS: 38  
; SEQ ID NO 4  
; LENGTH: 976  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
US-09-511-477-4

Query Match 100.0%; Score 167; DB 4; Length 976;  
Best Local Similarity 100.0%; Pred. No. 9.6e-17;  
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LFTVLDHNECLIIITGTEQYSEYTGVAETVR 31  
|||||

DB 770 LFTVLDHNECLIIITGTEQYSEYTGVAETVR 800  
|||||

RESULT 9  
US-09-511-507-4  
; Sequence 4, Application US/09511507  
; Patent No. 6395543  
; GENERAL INFORMATION:  
; APPLICANT: JACOBSON, Myron K.  
; APPLICANT: JACOBSON, Elaine L.  
; APPLICANT: AM, Jean-Christophe  
; APPLICANT: LIN, Winston  
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE (PARG) ENZ  
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIVE  
; TITLE OF INVENTION: THEREWITH  
; FILE REFERENCE: NIAD 201  
; CURRENT APPLICATION NUMBER: US/09/511,507  
; CURRENT FILING DATE: 2000-02-23  
; PRIOR APPLICATION NUMBER: 09/302,812  
; PRIOR FILING DATE: 1999-04-30  
; NUMBER OF SEQ ID NOS: 38  
; SEQ ID NO 4  
; LENGTH: 976  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
US-09-511-507-4

Query Match 100.0%; Score 167; DB 4; Length 976;  
Best Local Similarity 100.0%; Pred. No. 9.6e-17;  
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LFTVLDHNECLIIITGTEQYSEYTGVAETVR 31  
|||||  
DB 770 LFTVLDHNECLIIITGTEQYSEYTGVAETVR 800  
|||||

RESULT 10  
US-09-302-812-2  
; Sequence 2, Application US/09302812B  
; Patent No. 6333148  
; GENERAL INFORMATION:  
; APPLICANT: JACOBSON, Myron K.  
; APPLICANT: JACOBSON, Elaine L.  
; APPLICANT: AM, Jean-Christophe  
; APPLICANT: LIN, Winston  
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE (PARG) ENZ  
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIVE  
; TITLE OF INVENTION: THEREWITH  
; FILE REFERENCE: NIAD 201  
; CURRENT APPLICATION NUMBER: US/09/302,812B  
; CURRENT FILING DATE: 1999-04-30  
; EARLIER APPLICATION NUMBER: 60/083,768  
; EARLIER FILING DATE: 1998-05-01  
; NUMBER OF SEQ ID NOS: 38  
; SEQ ID NO 2  
; LENGTH: 977  
; TYPE: PRT  
; ORGANISM: Bos taurus  
; FEATURE:  
US-09-302-812-2

Query Match 100.0%; Score 167; DB 4; Length 977;  
Best Local Similarity 100.0%; Pred. No. 9.6e-17;  
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LFTVLDHNECLIIITGTEQYSEYTGVAETVR 31  
|||||  
DB 771 LFTVLDHNECLIIITGTEQYSEYTGVAETVR 801  
|||||

RESULT 11  
US-09-511-477-2

	Sequence 2, Application US/09511477	
	Patent No. 6337202	
	GENERAL INFORMATION:	
	APPLICANT: JACOBSON, Myron K.	
	APPLICANT: JACOBSON, Elaine L.	
	APPLICANT: AME, Jean-Christophe	
	APPLICANT: LIN, Winston	
	TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOSYLASE (PARG) EN	
	TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV	
	TITLE OF INVENTION: THEREWITH	
	FILE REFERENCE: NIAD 201	
	CURRENT APPLICATION NUMBER: US/09/511,477	
	CURRENT FILING DATE: 2000-02-23	
	PRIOR APPLICATION NUMBER: 09/302,812	
	PRIOR FILING DATE: 1999-04-30	
	NUMBER OF SEQ ID NOS: 38	
	SEQ ID NO 2	
	LENGTH: 977	
	TYPE: PRT	
	ORGANISM: Bos taurus	
	FEATURE:	
	US-09-511-477-2	
	Query Match	100.0%; Score 167; DB 4; Length 977;
	Best Local Similarity	100.0%; Pred. No. 9.6e-17;
	Matches	31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy	1 LFTEVLHDNECLIIITGTEQYSEYTGAEYR 31	
Db	771 LFTEVLHDNECLIIITGTEQYSEYTGAEYR 801	
	RESULT 12	
	US-09-511-507-2	
	Sequence 2, Application US/09511507	
	Patent No. 6395543	
	GENERAL INFORMATION:	
	APPLICANT: JACOBSON, Myron K.	
	APPLICANT: JACOBSON, Elaine L.	
	APPLICANT: AME, Jean-Christophe	
	APPLICANT: LIN, Winston	
	TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOSYLASE (PARG) EN	
	TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV	
	TITLE OF INVENTION: THEREWITH	
	FILE REFERENCE: NIAD 201	
	CURRENT APPLICATION NUMBER: US/09/511,507	
	CURRENT FILING DATE: 2000-02-23	
	PRIOR APPLICATION NUMBER: 09/302,812	
	PRIOR FILING DATE: 1999-04-30	
	NUMBER OF SEQ ID NOS: 38	
	SEQ ID NO 2	
	LENGTH: 977	
	TYPE: PRT	
	ORGANISM: Bos taurus	
	FEATURE:	
	US-09-511-507-2	
	Query Match	100.0%; Score 167; DB 4; Length 977;
	Best Local Similarity	100.0%; Pred. No. 9.6e-17;
	Matches	31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy	1 LFTEVLHDNECLIIITGTEQYSEYTGAEYR 31	
Db	771 LFTEVLHDNECLIIITGTEQYSEYTGAEYR 801	
	RESULT 13	
	US-09-302-812-8	
	Sequence 8, Application US/09302812B	
	Patent No. 6333148	
	GENERAL INFORMATION:	
	APPLICANT: JACOBSON, Myron K.	
	APPLICANT: JACOBSON, Elaine L.	

	APPLICANT: AM, Jean-Christophe	
	APPLICANT: LIN, Winston	
	TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOSYLASE (PARG) EN	
	TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV	
	TITLE OF INVENTION: THEREWITH	
	FILE REFERENCE: NIAD 201	
	CURRENT APPLICATION NUMBER: US/09/302,812B	
	CURRENT FILING DATE: 1999-04-30	
	EARLIER APPLICATION NUMBER: 60/083,768	
	EARLIER FILING DATE: 1998-05-01	
	NUMBER OF SEQ ID NOS: 38	
	SEQ ID NO 8	
	LENGTH: 768	
	TYPE: PRT	
	ORGANISM: Drosophila melanogaster	
	FEATURE:	
	US-09-302-812-8	
	Query Match	49.1%; Score 82; DB 4; Length 768;
	Best Local Similarity	53.3%; Pred. No. 0.00051;
	Matches	16; Conservative 5; Mismatches 9; Indels 0; Gaps 0;
Qy	1 LFTEVLHDNECLIIITGTEQYSEYTGAEY 30	
Db	400 LFTECLRPFPEALVMLGAERYSNYTGYAGSF 429	
	RESULT 14	
	US-09-511-477-8	
	Sequence 8, Application US/09511477	
	Patent No. 6337202	
	GENERAL INFORMATION:	
	APPLICANT: JACOBSON, Myron K.	
	APPLICANT: JACOBSON, Elaine L.	
	APPLICANT: AME, Jean-Christophe	
	APPLICANT: LIN, Winston	
	TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOSYLASE (PARG) EN	
	TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV	
	TITLE OF INVENTION: THEREWITH	
	FILE REFERENCE: NIAD 201	
	CURRENT APPLICATION NUMBER: US/09/511,477	
	CURRENT FILING DATE: 2000-02-23	
	PRIOR APPLICATION NUMBER: 09/302,812	
	PRIOR FILING DATE: 1999-04-30	
	NUMBER OF SEQ ID NOS: 38	
	SEQ ID NO 8	
	LENGTH: 768	
	TYPE: PRT	
	ORGANISM: Drosophila melanogaster	
	FEATURE:	
	US-09-511-477-8	
	Query Match	49.1%; Score 82; DB 4; Length 768;
	Best Local Similarity	53.3%; Pred. No. 0.00051;
	Matches	16; Conservative 5; Mismatches 9; Indels 0; Gaps 0;
Qy	1 LFTEVLHDNECLIIITGTEQYSEYTGAEY 30	
Db	400 LFTECLRPFPEALVMLGAERYSNYTGYAGSF 429	
	RESULT 15	
	US-09-511-507-8	
	Sequence 8, Application US/09511507	
	Patent No. 6395543	
	GENERAL INFORMATION:	
	APPLICANT: JACOBSON, Myron K.	
	APPLICANT: JACOBSON, Elaine L.	
	APPLICANT: AME, Jean-Christophe	
	APPLICANT: LIN, Winston	
	TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOSYLASE (PARG) EN	
	TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV	
	TITLE OF INVENTION: THEREWITH	

```
; FILE REFERENCE: NIAD 201
; CURRENT APPLICATION NUMBER: US/09/511.507
; CURRENT FILING DATE: 2000-02-23
; PRIOR APPLICATION NUMBER: 09/302,812
; PRIOR FILING DATE: 1999-04-30
; NUMBER OF SEQ ID NOS: 38
; SEQ ID NO 8
; LENGTH: 768
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
; FEATURE:
US-09-511-507-8

Query Match      49.1%; Score 82; DB 4; Length 768;
Best Local Similarity 53.3%; Pred. NO. 0.00051;
Matches 16; Conservative 5; Mismatches 9; Indels 0; Gaps 0;

QY      1 LFTEVLHDHNECLITGTQYSEYTGAAET 30
      ||||| | | | | | | | | | | | | | |
Db      400 LFTECLRFEPALVGLGARYSNITGYAGSF 429

Search completed: May 26, 2004, 18:49:30
Job time : 0.993349 secs
```

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 26, 2004, 17:19:25 ; Search time 3.55169 Seconds  
(without alignments)

2466.140 Million cell updates/sec

Title: US-09-302-812-11

Perfect score: 167

Sequence: 1 LFTEVLHDNECLIIINGTQSEYTGVAETYSR 31

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq\_29Jan04.\*

1: Geneseq1980s.\*

2: Geneseq1990s.\*

3: Geneseq2000s.\*

4: Geneseq2001s.\*

5: Geneseq2002s.\*

6: Geneseq2003as.\*

7: Geneseq2003bs.\*

8: Geneseq2004s.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	167	100.0	31	5 AAE25634	Aae25634 Bovine po
2	167	100.0	31	5 AAE25651	Aae25651 Bovine po
3	167	100.0	31	5 Aau76025	Aau76025 Bovine po
4	167	100.0	31	5 Aau76016	Aau76016 Bovine po
5	167	100.0	31	6 ABG72283	Abg72283 Oligopept
6	167	100.0	968	5 AAE25631	Aae25631 Murine po
7	167	100.0	968	5 Aau76022	Aau76022 Mouse pol
8	167	100.0	968	5 Aau76013	Aau76013 Mouse pol
9	167	100.0	968	5 ABG72280	Abg72280 Murine p
10	167	100.0	976	5 AAE25630	Aae25630 Human pol
11	167	100.0	976	5 Aau76021	Aau76021 Human pol
12	167	100.0	976	5 Aau76012	Aau76012 Human pol
13	167	100.0	976	6 ABG72279	Abg72279 Human pol
14	167	100.0	977	5 AAE25629	Aae25629 Bovine po
15	167	100.0	977	5 Aau76020	Aau76020 Bovine po
16	167	100.0	977	5 Aau75799	Aau75799 Bovine po
17	167	100.0	977	6 ABG72278	Abg72278 Bovine po
18	97	58.1	819	4 ABG20721	Abg20721 Novel hum
19	82	49.1	768	4 ABB59491	Abb59491 Drosophil
20	82	49.1	768	5 AAE25632	Aae25632 Fruit fly
21	82	49.1	768	5 Aau76023	Aau76023 Fruit fly
22	82	49.1	768	5 Aau76014	Aau76014 Fruit fly
23	82	49.1	768	6 ABG72281	Abg72281 Fruit fly
24	54.5	32.6	428	6 ABU24640	Abu24640 Protein e
25	53	31.7	726	5 AAE25633	Aae25633 Poly aden

26	53	31.7	726	5 AAU76024	Aau76024 Worm poly
27	53	31.7	726	5 AAU76015	Aau76015 Worm poly
28	52	31.7	726	6 ABG72282	Abg72282 C. elegans
29	52	31.1	464	4 AAB86542	Aab86542 E. rhaupon
30	50	29.9	172	1 AAP93508	Aap93508 Sequence
31	50	29.9	259	3 AAG16762	Aag16762 Arabidops
32	49.5	29.6	123	7 ADC95672	Adc95672 E. faeciu
33	49.5	29.6	307	6 ABU44051	Abu44051 Protein e
34	49.5	29.6	434	7 ADC94448	Adc94448 E. faeciu
35	49	29.3	447	6 ABU44183	Abu44183 Protein e
36	49	29.3	453	2 AAR86293	Aar86293 Protamino
37	49	29.3	3138	4 ABB61958	Abb61958 Drosophil
38	48.5	29.0	101	5 ABB10079	Abb10079 Human Hly
39	48.5	29.0	122	7 ADC96224	Adc96224 E. faeciu
40	48	28.7	372	2 AAW22175	Aaw22175 S.thermop
41	48	28.7	384	2 AAW14076	Aaw14076 S.thermop
42	48	28.7	993	5 AAE17835	Aae17835 Herpes si
43	48	28.7	1037	5 AAE17837	Aae17837 Herpes si
44	48	28.7	1057	2 AAW72067	Aaw72067 HSV-2 str
45	48	28.7	1113	5 AAE17836	Aae17836 Herpes si

## ALIGNMENTS

### RESULT 1

AAE25634

ID AAE25634 standard; peptide; 31 AA.

XX

AC AAE25634;

XX

DT 04-NOV-2002 (first entry)

XX

DE Bovine poly adenosine diphosphate-ribose glycohydrolase oligopeptide #1.

XX

KW Bovine; poly adenosine diphosphate-ribose glycohydrolase; bPARG; enzyme;

KW ADP-ribose; neoplastic disorder; genotoxic oxidative stress; neotropic;

KW cardiac disorder; neuronal disorder; reperfusion injury; neurotoxicity;

KW Alzheimer's disease; Huntington's disease; Parkinson's disease; cardiant;

KW cytoetic; vasotropic; neuroprotective; anticonvulsant; gene therapy;

KW antisense therapy.

XX

OS Bos taurus.

XX

PN US6395543-B1.

XX

PD 28-MAY-2002.

XX

PF 23-FEB-2000; 2000US-00511507.

XX

PR 01-MAY-1998; 98US-0083768P.

XX

PR 30-APR-1999; 99US-00302812.

XX

PA (KENT ) UNIV KENTUCKY RES FOUND.

XX

PI Jacobson MK, Jacobson EL, Ame J, Lin W;

XX

DR WPI; 2002-535641/57.

XX

XX New nucleic acid molecule encoding bovine poly adenosine diphosphate-

PT ribose glycohydrolase involved in cellular response to DNA damage,

PT inhibition of which is useful for treating neoplastic disorders and

PT neurodegenerative diseases.

XX

PS Claim 10; Col 81-82; 77pp; English.

XX

CC The invention relates to an isolated nucleic acid molecule which encodes

CC a bovine poly (adenosine diphosphate (ADP)-ribose) glycohydrolase (bPARG)

CC which catalyses release of ADP-ribose from an ADP-ribose polymer. The

CC invention is useful as probes and primer molecules that can be used in

CC hybridisation assays and polymerase chain reaction (PCR) amplification.

CC The knowledge of the nucleotide sequence of the PARG gene permits the

CC preparation of antisense therapeutics containing sequences complementary

CC to the mRNA of PARG gene. The antisense therapeutic are useful to treat  
 CC neoplastic disorders and conditions caused by genotoxic oxidative stress  
 CC e.g., cardiac disorders, neuronal disorders, reperfusion injury,  
 CC neurotoxicity, Alzheimer's disease, Huntington's disease and Parkinson's  
 CC disease. The invention is useful in gene therapy and antisense therapy.  
 CC The present sequence is bPARG oligopeptide  
 XX  
 SQ Sequence 31 AA;  
 Query Match 100.0%; Score 167; DB 5; Length 31;  
 Best Local Similarity 100.0%; Pred. No. 6.8e-17;  
 Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 LFTFVLHDNECLIIITGTEQYSEYTGVAETVR 31  
 DB 1 LFTFVLHDNECLIIITGTEQYSEYTGVAETVR 31  
 RESULT 2  
 ID AAE25651 standard; peptide; 31 AA.  
 AC AAE25651;  
 DT 04-NOV-2002 (first entry)  
 DE Bovine poly adenosine diphosphate-ribose glycohydrolase oligopeptide #7.  
 KW Bovine; poly adenosine diphosphate-ribose glycohydrolase; bPARG; enzyme;  
 KW ADP-ribose; neoplastic disorder; genotoxic oxidative stress; nootropic;  
 KW cardiac disorder; neuronal disorder; reperfusion injury; neurotoxicity;  
 KW Alzheimer's disease; Huntington's disease; Parkinson's disease; cardiac;  
 KW cytosolic; vasotropic; neuroprotective; anticonvulsant; gene therapy;  
 KW antisense therapy.  
 OS Bos taurus.  
 XX USG395543-B1.  
 FN 28-MAY-2002.  
 PD 23-FEB-2000; 2000US-00511507.  
 PF 01-MAY-1998; 98US-0083768P.  
 PR 30-APR-1999; 99US-00302812.  
 XX (KENT ) UNIV KENTUCKY RES FOUND.  
 PA Jacobson MK, Jacobson EL, Ame J, Lin W;  
 PI WPI; 2002-535641/57.  
 DR New nucleic acid molecule encoding bovine poly adenosine diphosphate-  
 PT ribose glycohydrolase involved in cellular response to DNA damage,  
 PT inhibition of which is useful for treating neoplastic disorders and  
 PT neurodegenerative diseases.  
 XX Example 3; Col 27; 77pp; English.  
 PS The invention relates to an isolated nucleic acid molecule which encodes  
 CC a bovine poly (adenosine diphosphate (ADP)-ribose) glycohydrolase (bPARG)  
 CC which catalyses release of ADP-ribose from an ADP ribose polymer. The  
 CC invention is useful as probes and primer molecules that can be used in  
 CC hybridisation assays and polymerase chain reaction (PCR) amplification.  
 CC The knowledge of the nucleotide sequence of the PARG gene permits the  
 CC preparation of antisense therapeutics containing sequences complementary  
 CC to the mRNA of PARG gene. The antisense therapeutic are useful to treat  
 CC neoplastic disorders and conditions caused by genotoxic oxidative stress  
 CC e.g., cardiac disorders, neuronal disorders, reperfusion injury,  
 CC neurotoxicity, Alzheimer's disease, Huntington's disease and Parkinson's  
 CC disease. The invention is useful in gene therapy and antisense therapy.  
 CC The present sequence is bPARG oligopeptide  
 XX

SQ Sequence 31 AA;  
 Query Match 100.0%; Score 167; DB 5; Length 31;  
 Best Local Similarity 100.0%; Pred. No. 6.8e-17;  
 Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 LFTFVLHDNECLIIITGTEQYSEYTGVAETVR 31  
 DB 1 LFTFVLHDNECLIIITGTEQYSEYTGVAETVR 31  
 RESULT 3  
 ID AAU76025 standard; peptide; 31 AA.  
 AC AAU76025;  
 DT 08-MAY-2002 (first entry)  
 DE Bovine poly (ADP-ribose) glycohydrolase (PARG) oligopeptide 68.  
 KW Bovine; poly (ADP-ribose) glycohydrolase; PARG; DNA repair; ADP-ribose;  
 KW adenosine diphosphate-ribose; neoplastic disorder; adenocarcinoma;  
 KW leukaemia; lymphoma; melanoma; myeloma; sarcoma; teratocarcinoma;  
 KW hyperplasia; hypertrophy; reperfusion; ischaemia; heart attack; stroke;  
 KW neurodegenerative disease; neurological disorder; Alzheimer's disease;  
 KW Huntington's disease; Parkinson's disease; oligopeptide 68.  
 OS Bos taurus.  
 XX USG337202-B1.  
 FN 08-JAN-2002.  
 PD 23-FEB-2000; 2000US-00511477.  
 PF 01-MAY-1998; 98US-0083768P.  
 PR 30-APR-1999; 99US-00302812.  
 XX (KENT ) UNIV KENTUCKY RES FOUND.  
 PA Jacobson MK, Jacobson EL, Ame J, Lin W;  
 PI WPI; 2002-163240/21.  
 DR Novel isolated and purified poly (ADP-ribose) glycohydrolase protein  
 PT which catalyses release of ADP-ribose from ADP ribose polymer, useful for  
 PT treating neoplastic and neurological disorders, heart attack and stroke.  
 XX Claim 4; Col 25; 81pp; English.  
 PS The present invention relates to a new poly (ADP-ribose) glycohydrolase  
 CC (PARG) protein which catalyses release of ADP-ribose from an ADP  
 CC (adenosine diphosphate)-ribose polymer. The PARG molecule of the  
 CC invention is useful for generating antibodies and can be inhibited or  
 CC activated for diagnosing and treating neoplastic disorders such as  
 CC adenocarcinoma, leukaemia, lymphoma, melanoma, myeloma, sarcoma,  
 CC teratocarcinoma, hyperplasia and hypertrophy, reperfusion following  
 CC ischaemia, heart attack, stroke, neurodegenerative diseases, neurological  
 CC disorders including Alzheimer's, Huntington's and Parkinson's diseases,  
 CC and related conditions. PARG levels may be enhanced to suppress DNA  
 CC repair and increase the cell's susceptibility to chemotherapy drugs.  
 CC Antagonists of PARG are administered to treat or prevent neoplastic  
 CC disorders. The present amino acid sequence represents bovine PARG  
 CC oligonucleotide 68. This peptide is one of several PARG oligopeptides  
 CC (AAU76025-AAU76028) of the invention  
 XX  
 SQ Sequence 31 AA;  
 Query Match 100.0%; Score 167; DB 5; Length 31;  
 Best Local Similarity 100.0%; Pred. No. 6.8e-17;  
 Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LFTFVLHDNECLIIITGTEQYSEYTGVAETVR 31  
 DB 1 LFTFVLHDNECLIIITGTEQYSEYTGVAETVR 31

RESULT 4  
 AAU76016  
 ID AAU76016 standard; peptide; 31 AA.  
 AC AAU76016;  
 XX  
 DT 08-MAY-2002 (first entry)  
 XX  
 DE Bos taurus.  
 XX  
 XX Bovine poly(ADP-ribose) glycohydrolase (PARG) oligopeptide 68.  
 XX  
 XX Cow; poly(ADP-ribose) glycohydrolase; PARG; PARG modulator; ADP-ribose;  
 KW adenosine diphosphate-ribose; DNA repair; apoptosis; neoplasia; cancer;  
 KW inherited genetic disease; myocardial infarction; vascular stroke; aging;  
 KW neurodegeneration; Huntington's disease; Parkinson's disease;  
 KW Alzheimer's disease; neurotoxicity; oligopeptide 68.  
 XX  
 OS Bos taurus.  
 XX  
 XX US6333148-B1.  
 XX  
 XX 25-DEC-2001.  
 XX  
 XX 30-APR-1999; 99US-00302812.  
 XX  
 XX 01-MAY-1998; 98US-0083768P.  
 XX  
 XX (KENT) UNIV KENTUCKY RES FOUND.  
 XX  
 XX Jacobson MK, Jacobson EL, Ame J, Lin W;  
 PI WPI; 2002-153820/20.  
 XX  
 XX Screening compounds for modulation of poly(ADP-ribose) glycohydrolase,  
 PT useful potentially for treating diseases associated with DNA damage, e.g.  
 PT cancer.  
 PT  
 XX Claim 4; Col 25; 80pp; English.  
 XX  
 XX The present invention relates to a new method for screening compounds for  
 CC ability to modulate activity of an enzyme that hydrolyses ADP (adenosine  
 CC diphosphate)-ribose from an ADP-ribose polymer. The compounds are  
 CC inhibitors or activators of PARG (poly(ADP-ribose) glycohydrolase) and  
 CC are used to treat or prevent any condition associated with DNA damage,  
 CC e.g. neoplasia, inherited genetic diseases, myocardial infarction,  
 CC vascular stroke, aging and neurodegeneration e.g. Huntington's,  
 CC Parkinson's or Alzheimer's diseases, or neurotoxicity generally.  
 CC Compounds identified by the new method are more effective than known  
 CC inhibitors and have fewer side effects. The present amino acid sequence  
 CC represents bovine PARG oligopeptide 68. This peptide is one of several  
 CC PARG oligopeptides (AAU76016-AAU76019) of the invention  
 XX  
 XX Sequence 31 AA;  
 XX  
 XX Query Match 100.0%; Score 167; DB 5; Length 31;  
 XX Best Local Similarity 100.0%; Pred. No. 6.8e-17;  
 XX Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 XX

QY 1 LFTFVLHDNECLIIITGTEQYSEYTGVAETVR 31  
 DB 1 LFTFVLHDNECLIIITGTEQYSEYTGVAETVR 31

RESULT 5  
 ABG72283  
 ID ABG72283 standard; peptide; 31 AA.  
 XX  
 AC ABG72283;  
 XX

DT 13-MAR-2003 (first entry)  
 XX  
 DE Oligopeptide #1 derived from bovine PARG enzyme.  
 XX  
 XX Bovine; poly(ADP-ribose) glycohydrolase; PARG; PARG expression;  
 KW cellular response; DNA damage; neoplastic disorder inducing agent;  
 KW oxidative stress; neoplastic disorder; myocardial infarction;  
 KW vascular stroke; neurodegenerative disorder; Alzheimer's disease;  
 KW Parkinson's disease; Huntington's disease; inborn genetic error;  
 KW reperfusion; ischaemia; aging; neurotoxicity; PARG activity; cytostatic;  
 KW neuroprotective; neurotropic; antiparkinsonian; cardiant; vasotropic;  
 KW anticonvulsant; cerebroprotective.  
 XX  
 OS Bos taurus.  
 XX  
 XX US2002132328-A1.  
 XX  
 XX 19-SEP-2002.  
 XX  
 XX 09-OCT-2001; 2001US-00973451.  
 XX  
 XX 01-MAY-1998; 98US-0083768P.  
 XX  
 XX 30-APR-1999; 99US-00302812.  
 XX  
 XX (JACO/) JACOBSON M K.  
 XX (JACO/) JACOBSON E L.  
 XX (AMEJ/) AME J.  
 XX (LINW/) LIN W.  
 XX  
 XX Jacobson MK, Jacobson EL, Ame J, Lin W;  
 XX WPI; 2003-155895/15.  
 XX  
 XX New nucleic acid molecule encoding a polypeptide with poly(ADP-ribose)  
 PT glycohydrolase activity, for preventing, treating, or ameliorating a  
 PT disease condition, e.g. neoplastic disorder, myocardial infarction or  
 PT vascular stroke.  
 XX  
 XX Example 2; Page 14; 86pp; English.  
 XX  
 XX The present invention relates to the isolation of poly(ADP-ribose)  
 CC glycohydrolase (PARG) from several species, and the polynucleotide  
 CC sequences encoding them. Methods for inhibiting PARG expression or  
 CC overexpressing PARG are also disclosed. PARG is involved in the cellular  
 CC response to DNA damage, and is associated with the body's response to  
 CC neoplastic disorder inducing agents and oxidative stress. The  
 CC polynucleotide sequences encoding PARG and PARG modulators are useful for  
 CC preventing, treating, or ameliorating diseases such as neoplastic  
 CC disorders, myocardial infarction, vascular stroke, neurodegenerative  
 CC diseases (e.g. Alzheimer's disease, Parkinson's disease or Huntington's  
 CC disease), inborn genetic errors, reperfusion following ischaemia, aging,  
 CC and neurotoxicity. The polynucleotide sequences are also useful in gene  
 CC therapy. The methods are useful for identifying an agent that modulates  
 CC PARG activity, identifying a mutant PARG allele in an individual, or  
 CC screening candidate molecules for PARG modulating activity. ABG72283-  
 CC ABG72286 represent oligopeptides derived from bovine PARG enzyme. The  
 CC oligopeptides are used to construct degenerate PCR primers for the  
 CC isolation of cDNA encoding bovine PARG  
 XX  
 XX Sequence 31 AA;  
 XX  
 XX Query Match 100.0%; Score 167; DB 6; Length 31;  
 XX Best Local Similarity 100.0%; Pred. No. 6.8e-17;  
 XX Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 XX

QY 1 LFTFVLHDNECLIIITGTEQYSEYTGVAETVR 31  
 DB 1 LFTFVLHDNECLIIITGTEQYSEYTGVAETVR 31

RESULT 6  
 AAE25631  
 ID AAE25631 standard; protein; 968 AA.



XX AC AAE25631;  
XX DT 04-NOV-2002 (first entry)  
XX DE Murine poly adenosine diphosphate-ribose glycohydrolase (PARG).  
XX  
XX Murine; poly adenosine diphosphate-ribose glycohydrolase; PARG; enzyme;  
KW ADP-ribose; neoplastic disorder; genotoxic oxidative stress; neurotropic;  
KW cardiac disorder; neuronal disorder; reperfusion injury; neurotoxicity;  
KW Alzheimer's disease; Huntington's disease; Parkinson's disease; cardiant;  
KW cytotatic; vasotropic; neuroprotective; anticonvulsant; gene therapy;  
KW antisense therapy.  
XX  
XX Mus musculus.  
XX  
XX US6395543-B1.  
XX PN  
XX 28-MAY-2002.  
XX PD  
XX 23-FEB-2000; 2000US-005111507.  
XX PF  
XX 01-MAY-1998; 98US-0083768P.  
XX PR  
XX 30-APR-1999; 99US-00302812.  
XX  
XX (KENT ) UNIV KENTUCKY RES FOUND.  
XX PA  
XX Jacobson MK, Jacobson EL, Ame J, Lin W;  
XX WPI; 2002-535641/57.  
XX DR N-PSDB; AAD42083.  
XX  
XX New nucleic acid molecule encoding bovine poly adenosine diphosphate-  
PT ribose glycohydrolase involved in cellular response to DNA damage,  
PT inhibition of which is useful for treating neoplastic disorders and  
PT neurodegenerative diseases.  
XX  
XX Claim 3; Col 63-68; 77pp; English.  
XX  
XX The invention relates to an isolated nucleic acid molecule which encodes  
CC a bovine poly (adenosine diphosphate (ADP)-ribose) glycohydrolase (bPARG)  
CC which catalyses release of ADP-ribose from an ADP ribose polymer. The  
CC invention is useful as probes and primer molecules that can be used in  
CC hybridisation assays and polymerase chain reaction (PCR) amplification.  
CC The knowledge of the nucleotide sequence of the PARG gene permits the  
CC preparation of antisense therapeutics containing sequences complementary  
CC to the mRNA of PARG gene. The antisense therapeutic are useful to treat  
CC neoplastic disorders and conditions caused by genotoxic oxidative stress  
CC e.g., cardiac disorders, neuronal disorders, reperfusion injury,  
CC neurotoxicity, Alzheimer's disease, Huntington's disease and Parkinson's  
CC disease. The invention is useful in gene therapy and antisense therapy.  
XX  
XX The present sequence is murine PARG  
XX  
SQ Sequence 968 AA;  
Query Match 100.0%; Score 167; DB 5; Length 968;  
Best Local Similarity 100.0%; Pred. No. 3.9e-15;  
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 LFTVELDHNECLIIITGTQYSEYTGVAETVR 31  
DB 762 LFTVELDHNECLIIITGTQYSEYTGVAETVR 792  
RESULT 7  
AAU76022  
ID AAU76022 standard; protein; 968 AA.  
XX  
XX AAU76022;  
XX  
XX 08-MAY-2002 (first entry)  
XX  
XX Mouse poly (ADP-ribose) glycohydrolase (PARG) protein sequence.

XX KW Mouse; poly (ADP-ribose) glycohydrolase; PARG; DNA repair; ADP-ribose;  
KW adenosine diphosphate-ribose; neoplastic disorder; adenocarcinoma;  
KW leukaemia; lymphoma; melanoma; myeloma; sarcoma; teratocarcinoma;  
KW hyperplasia; hypertrophy; reperfusion; ischaemia; heart attack; stroke;  
KW neurodegenerative disease; neurological disorder; Alzheimer's disease;  
KW Huntington's disease; Parkinson's disease.  
XX  
XX Mus musculus.  
XX  
XX US6337202-B1.  
XX PN  
XX 08-JAN-2002.  
XX PD  
XX 23-FEB-2000; 2000US-00511477.  
XX PF  
XX 01-MAY-1998; 98US-0083768P.  
XX PR  
XX 30-APR-1999; 99US-00302812.  
XX  
XX (KENT ) UNIV KENTUCKY RES FOUND.  
XX PA  
XX Jacobson MK, Jacobson EL, Ame J, Lin W;  
XX WPI; 2002-163240/21.  
XX DR N-PSDB; ABK14933.  
XX  
XX Novel isolated and purified poly (ADP-ribose) glycohydrolase protein  
PT which catalyses release of ADP-ribose from ADP ribose polymer, useful for  
PT treating neoplastic and neurological disorders, heart attack and stroke.  
XX  
XX Claim 2; Col 63-70; 81pp; English.  
XX  
XX The present invention relates to a new poly (ADP-ribose) glycohydrolase  
CC (PARG) protein which catalyses release of ADP-ribose from an ADP  
CC (adenosine diphosphate)-ribose polymer. The PARG molecule of the  
CC invention is useful for generating antibodies and can be inhibited or  
CC activated for diagnosing and treating neoplastic disorders such as  
CC adenocarcinoma, leukaemia, lymphoma, melanoma, myeloma, sarcoma,  
CC teratocarcinoma, hyperplasia and hypertrophy, reperfusion following  
CC ischaemia, heart attack, stroke, neurodegenerative diseases, neurological  
CC disorders including Alzheimer's, Huntington's and Parkinson's diseases,  
CC and related conditions. PARG levels may be enhanced to suppress DNA  
CC repair and increase the cell's susceptibility to chemotherapy drugs.  
CC Antagonists of PARG are administered to treat or prevent neoplastic  
CC disorders. The present amino acid sequence represents the mouse PARG  
CC protein of the invention. This protein is one of several PARG proteins  
CC (AAU76020-AAU76024) of the invention  
XX  
SQ Sequence 968 AA;  
Query Match 100.0%; Score 167; DB 5; Length 968;  
Best Local Similarity 100.0%; Pred. No. 3.9e-15;  
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 LFTVELDHNECLIIITGTQYSEYTGVAETVR 31  
DB 762 LFTVELDHNECLIIITGTQYSEYTGVAETVR 792  
RESULT 8  
AAU76013  
ID AAU76013 standard; protein; 968 AA.  
XX  
XX AAU76013;  
XX  
XX 08-MAY-2002 (first entry)  
XX  
XX Mouse poly (ADP-ribose) glycohydrolase (PARG) protein sequence.  
XX  
XX Mouse; poly (ADP-ribose) glycohydrolase; PARG; PARG modulator; ADP-ribose;  
KW adenosine diphosphate-ribose; DNA repair; apoptosis; neoplasia; cancer;  
KW inherited genetic disease; myocardial infarction; vascular stroke; aging;  
KW neurodegeneration; Huntington's disease; Parkinson's disease;

KW Alzheimer's disease; neurotoxicity.  
 XX Mus musculus.  
 OS US6333148-B1.  
 XX  
 PN  
 XX  
 XX 25-DEC-2001.  
 XX 30-APR-1999; 99US-00302812.  
 XX 01-MAY-1998; 98US-0083768P.  
 PR (KENT ) UNIV KENTUCKY RES FOUND.  
 XX Jacobson MK, Jacobson EL, Ame J, Lin W;  
 XX WPI; 2002-153820/20.  
 XX N-PSDB; ABK14495.  
 DR Screening compounds for modulation of poly(ADP-ribose) glycohydrolase,  
 XX useful potentially for treating diseases associated with DNA damage, e.g.  
 PT cancer.  
 XX  
 XX Claim 3; Col 63-68; 80pp; English.  
 XX The present invention relates to a new method for screening compounds for  
 CC ability to modulate activity of an enzyme that hydrolyses ADP (adenosine  
 CC diphosphate)-ribose from an ADP-ribose polymer. The compounds are  
 CC inhibitors or activators of PARG (poly(ADP-ribose) glycohydrolase) and  
 CC are used to treat or prevent any condition associated with DNA damage,  
 CC e.g. neoplasia, inherited genetic diseases, myocardial infarction,  
 CC vascular stroke, aging and neurodegeneration e.g. Huntington's,  
 CC Parkinson's or Alzheimer's diseases, or neurotoxicity generally.  
 CC Compounds identified by the new method are more effective than known  
 CC inhibitors and have fewer side effects. The present amino acid sequence  
 CC represents the mouse PARG protein of the invention. This protein is one  
 CC of several PARG proteins (AAU75799 and AAU76012-AAU76015) of the  
 CC invention  
 XX  
 XX Sequence 968 AA;  
 SQ  
 Query Match 100.0%; Score 167; DB 5; Length 968;  
 Best Local Similarity 100.0%; Pred. No. 3.9e-15;  
 Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 LFTFVLHDNECLITGTQYSEYTGVAETVR 31  
 Db 762 LFTFVLHDNECLITGTQYSEYTGVAETVR 792  
 RESULT 9  
 ABG72280  
 ID ABG72280 standard; protein; 968 AA.  
 XX  
 AC ABG72280;  
 XX  
 XX 13-MAR-2003 (first entry)  
 XX  
 XX Murine poly (ADP-ribose) glycohydrolase (PARG) enzyme.  
 DE  
 XX Murine; poly (ADP-ribose) glycohydrolase; PARG; PARG expression;  
 XX cellular response; DNA damage; neoplastic disorder inducing agent;  
 KW oxidative stress; neoplastic disorder; myocardial infarction;  
 KW vascular stroke; neurodegenerative disorder; Alzheimer's disease;  
 KW Parkinson's disease; Huntington's disease; inborn genetic error;  
 KW reperfusion; ischaemia; aging; neurotoxicity; PARG activity; mouse;  
 KW cytosolic; neuroprotective; neurotropic; antiparkinsonian; cardiant;  
 KW vasotropic; anticonvulsant; cerebroprotective; enzyme.  
 XX  
 OS Mus musculus.  
 XX  
 XX US2002132328-A1.  
 XX

PD 19-SEP-2002.  
 XX 09-OCT-2001; 2001US-00973451.  
 XX  
 XX 01-MAY-1998; 98US-0083768P.  
 PR 30-APR-1999; 99US-00302812.  
 XX (JACO/) JACOBSON M K.  
 PA (JACO/) JACOBSON E L.  
 PA (AMEJ/) AME J.  
 PA (LINW/) LIN W.  
 XX Jacobson MK, Jacobson EL, Ame J, Lin W;  
 XX WPI; 2003-155895/15.  
 DR N-PSDB; ABX14479.  
 XX New nucleic acid molecule encoding a polypeptide with poly(ADP-ribose)  
 PT glycohydrolase activity, for preventing, treating, or ameliorating a  
 PT disease condition, e.g. neoplastic disorder, myocardial infarction or  
 PT vascular stroke.  
 XX  
 XX Claim 28; Fig 16; 86pp; English.  
 XX The present invention relates to the isolation of poly(ADP-ribose)  
 CC glycohydrolase (PARG) from several species, and the polynucleotide  
 CC sequences encoding them. Methods for inhibiting PARG expression or  
 CC overexpressing PARG are also disclosed. PARG is involved in the cellular  
 CC response to DNA damage, and is associated with the body's response to  
 CC neoplastic disorder inducing agents and oxidative stress. The  
 CC polynucleotide sequences encoding PARG and PARG modulators are useful for  
 CC preventing, treating, or ameliorating diseases such as neoplastic  
 CC disorders, myocardial infarction, vascular stroke, neurodegenerative  
 CC disorders (e.g. Alzheimer's disease, Parkinson's disease or Huntington's  
 CC disease), inborn genetic errors, reperfusion following ischaemia, aging,  
 CC and neurotoxicity. The polynucleotide sequences are also useful in gene  
 CC therapy. The methods are useful for identifying an agent that modulates  
 CC PARG activity, identifying a mutant PARG allele in an individual, or  
 CC screening candidate molecules for PARG modulating activity. The present  
 CC sequence represents murine PARG enzyme  
 XX  
 XX Sequence 968 AA;  
 SQ  
 Query Match 100.0%; Score 167; DB 6; Length 968;  
 Best Local Similarity 100.0%; Pred. No. 3.9e-15;  
 Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 LFTFVLHDNECLITGTQYSEYTGVAETVR 31  
 Db 762 LFTFVLHDNECLITGTQYSEYTGVAETVR 792  
 RESULT 10  
 AAE25630  
 ID AAE25630 standard; protein; 976 AA.  
 XX  
 XX AAE25630;  
 AC  
 XX 04-NOV-2002 (first entry)  
 XX  
 XX Human poly adenosine diphosphate-ribose glycohydrolase (PARG).  
 DE  
 XX Human; poly adenosine diphosphate-ribose glycohydrolase; PARG; enzyme;  
 KW ADP-ribose; neoplastic disorder; genotoxic oxidative stress; neurotropic;  
 KW cardiac disorder; neuronal disorder; reperfusion injury; neurotoxicity;  
 KW Alzheimer's disease; Huntington's disease; Parkinson's disease; cardiant;  
 KW cytosolic; vasotropic; neuroprotective; anticonvulsant; gene therapy;  
 KW antisense therapy.  
 XX  
 XX Homo sapiens.  
 OS  
 XX US6395543-B1.  
 XX PN  
 XX

PD 28-MAY-2002.  
 XX  
 PF 23-FEB-2000; 2000US-00511507.  
 XX  
 XX 01-MAY-1998; 98US-0083768P.  
 PR 30-APR-1999; 99US-00302812.  
 XX  
 XX (KENT ) UNIV KENTUCKY RES FOUND.  
 PA  
 XX Jacobson MK, Jacobson EL, Ame J, Lin W;  
 PI WPI; 2002-535641/57.  
 XX N-PSDB; AAD42082.  
 DR  
 XX New nucleic acid molecule encoding bovine poly adenosine diphosphate-  
 XX ribose glycohydrolase involved in cellular response to DNA damage.  
 PT  
 PT inhibition of which is useful for treating neoplastic disorders and  
 PT neurodegenerative diseases.  
 XX  
 XX Claim 3; Col 55-60; 77pp; English.  
 PS  
 XX The invention relates to an isolated nucleic acid molecule which encodes  
 XX a bovine poly (adenosine diphosphate (ADP)-ribose) glycohydrolase (bPARG)  
 CC which catalyses release of ADP-ribose from an ADP ribose polymer. The  
 CC invention is useful as probes and primer molecules that can be used in  
 CC hybridisation assays and polymerase chain reaction (PCR) amplification.  
 CC The knowledge of the nucleotide sequence of the PARG gene permits the  
 CC preparation of antisense therapeutics containing sequences complementary  
 CC to the mRNA of PARG gene. The antisense therapeutics are useful to treat  
 CC neoplastic disorders and conditions caused by genotoxic oxidative stress  
 CC e.g., cardiac disorders, neuronal disorders, reperfusion injury.  
 CC neurotoxicity, Alzheimer's disease, Huntington's disease and Parkinson's  
 CC disease. The invention is useful in gene therapy and antisense therapy.  
 CC The present sequence is human PARG  
 XX  
 XX Sequence 976 AA;  
 SQ  
 Query Match 100.0%; Score 167; DB 5; Length 976;  
 Best Local Similarity 100.0%; Pred. No. 3.9e-15;  
 Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Qy 1 LFTEVLHNECLIIITGTQYSEYTGATYR 31  
 Db 770 LFTEVLHNECLIIITGTQYSEYTGATYR 800  
 RESULT 11  
 AAU76021  
 ID AAU76021 standard; protein; 976 AA.  
 XX  
 AC AAU76021;  
 XX  
 XX 08-MAY-2002 (first entry)  
 DT  
 DE Human poly (ADP-ribose) glycohydrolase (PARG) protein sequence.  
 XX  
 XX Human; poly (ADP-ribose) glycohydrolase; PARG; DNA repair; ADP-ribose;  
 KW adenosine diphosphate-ribose; neoplastic disorder; adenocarcinoma;  
 KW leukaemia; lymphoma; melanoma; myeloma; sarcoma; teratocarcinoma;  
 KW hyperplasia; hypertrophy; reperfusion; ischaemia; heart attack; stroke;  
 KW neurodegenerative disease; neurological disorder; Alzheimer's disease;  
 KW Huntington's disease; Parkinson's disease.  
 XX  
 XX Homo sapiens.  
 OS  
 XX US6337202-B1.  
 PN  
 XX 08-JAN-2002.  
 PD  
 XX 23-FEB-2000; 2000US-00511477.  
 PF  
 XX 01-MAY-1998; 98US-0083768P.  
 PR 30-APR-1999; 99US-00302812.  
 PR

XX (KENT ) UNIV KENTUCKY RES FOUND.  
 PA  
 XX Jacobson MK, Jacobson EL, Ame J, Lin W;  
 PI WPI; 2002-153240/21.  
 XX N-PSDB; ABK14932.  
 DR  
 XX Novel isolated and purified poly (ADP-ribose) glycohydrolase protein  
 XX which catalyses release of ADP-ribose from ADP ribose polymer, useful for  
 PT treating neoplastic and neurological disorders, heart attack and stroke.  
 PT  
 XX Claim 2; Col 55-60; 81pp; English.  
 PS  
 XX The present invention relates to a new poly (ADP-ribose) glycohydrolase  
 CC (PARG) protein which catalyses release of ADP-ribose from an ADP  
 CC (adenosine diphosphate)-ribose polymer. The PARG molecule of the  
 CC invention is useful for generating antibodies and can be inhibited or  
 CC activated for diagnosing and treating neoplastic disorders such as  
 CC adenocarcinoma, leukaemia, lymphoma, melanoma, myeloma, sarcoma,  
 CC teratocarcinoma, hyperplasia and hypertrophy, reperfusion following  
 CC ischaemia, heart attack, stroke, neurodegenerative diseases, neurological  
 CC disorders including Alzheimer's, Huntington's and Parkinson's diseases,  
 CC and related conditions. PARG levels may be enhanced to suppress DNA  
 CC repair and increase the cell's susceptibility to chemotherapy drugs.  
 CC Antagonists of PARG are administered to treat or prevent neoplastic  
 CC disorders. The present amino acid sequence represents the human PARG  
 CC protein of the invention. This protein is one of several PARG proteins  
 CC (AAU76020-AAU76024) of the invention  
 XX  
 XX Sequence 976 AA;  
 SQ  
 Query Match 100.0%; Score 167; DB 5; Length 976;  
 Best Local Similarity 100.0%; Pred. No. 3.9e-15;  
 Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Qy 1 LFTEVLHNECLIIITGTQYSEYTGATYR 31  
 Db 770 LFTEVLHNECLIIITGTQYSEYTGATYR 800  
 RESULT 12  
 AAU76012  
 ID AAU76012 standard; protein; 976 AA.  
 XX  
 AC AAU76012;  
 XX  
 XX 08-MAY-2002 (first entry)  
 DT  
 DE Human poly (ADP-ribose) glycohydrolase (PARG) protein sequence.  
 XX  
 XX Human; poly (ADP-ribose) glycohydrolase; PARG; PARG modulator; ADP-ribose;  
 KW adenosine diphosphate-ribose; DNA repair; apoptosis; neoplasia; cancer;  
 KW inherited genetic disease; myocardial infarction; vascular stroke; aging;  
 KW neurodegeneration; Huntington's disease; Parkinson's disease;  
 KW Alzheimer's disease; neurotoxicity.  
 XX  
 XX Homo sapiens.  
 OS  
 XX US6333148-B1.  
 PN  
 XX 25-DEC-2001.  
 PD  
 XX 30-APR-1999; 99US-00302812.  
 PF  
 XX 01-MAY-1998; 98US-0083768P.  
 PR  
 XX (KENT ) UNIV KENTUCKY RES FOUND.  
 PA  
 XX Jacobson MK, Jacobson EL, Ame J, Lin W;  
 PI WPI; 2002-153820/20.  
 XX N-PSDB; ABK14494.  
 DR

XX Screening compounds for modulation of poly(ADP-ribose) glycohydrolase,  
PT useful potentially for treating diseases associated with DNA damage, e.g.  
PT cancer.  
XX  
XX Claim 3; Col 55-60; 80pp; English.  
XX  
XX The present invention relates to a new method for screening compounds for  
CC ability to modulate activity of an enzyme that hydrolyses ADP (adenosine  
CC diphosphate)-ribose from an ADP-ribose polymer. The compounds are  
CC inhibitors or activators of PARG (poly(ADP-ribose) glycohydrolase) and  
CC are used to treat or prevent any condition associated with DNA damage  
CC e.g. neoplasia, inherited genetic diseases, myocardial infarction,  
CC vascular stroke, aging and neurodegeneration e.g. Huntington's,  
CC Parkinson's or Alzheimer's diseases, or neurotoxicity generally.  
CC Compounds identified by the new method are more effective than known  
CC inhibitors and have fewer side effects. The present amino acid sequence  
CC represents the human PARG protein of the invention. This protein is one  
CC of several PARG proteins (AAU75799 and AAU76012-AAU76015) of the  
CC invention  
XX  
XX Sequence 976 AA;  
SQ  
Query Match 100.0%; Score 167; DB 5; Length 976;  
Best Local Similarity 100.0%; Pred. No. 3.9e-15;  
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 LFTVLDHNECLIIITGTOYSEYTGVAETYSR 31  
DB 770 LFTVLDHNECLIIITGTOYSEYTGVAETYSR 800  
  
XX  
XX RESULT 13  
XX ABG72279  
XX ID ABG72279 standard; protein; 976 AA.  
XX AC ABG72279;  
XX DT 13-MAR-2003 (first entry)  
XX DE Human poly(ADP-ribose) glycohydrolase (PARG) enzyme.  
XX  
XX Human; poly(ADP-ribose) glycohydrolase; PARG; PARG expression;  
XX cellular response; DNA damage; neoplastic disorder inducing agent;  
XX oxidative stress; neoplastic disorder; myocardial infarction;  
XX vascular stroke; neurodegenerative disorder; Alzheimer's disease;  
XX Parkinson's disease; Huntington's disease; inborn genetic error;  
XX reperfusion; ischaemia; aging; neurotoxicity; PARG activity; cytostatic;  
XX neuroprotective; neurotropic; antiparkinsonian; cardiatic; vasotropic;  
XX anticonvulsant; cerebroprotective; enzyme.  
XX  
XX Homo sapiens.  
XX OS  
XX US2002132328-A1.  
XX PN  
XX 19-SEP-2002.  
XX PD  
XX  
XX 09-OCT-2001; 2001US-00973451.  
XX PF  
XX 01-MAY-1998; 98US-0083768P.  
XX PR  
XX 30-APR-1999; 99US-00302812.  
XX PR  
XX (JACO/) JACOBSON M K.  
XX PA (JACO/) JACOBSON E L.  
XX PA (AMEJ/) AME J.  
XX PA (LINW/) LIN W.  
XX  
XX Jacobson MK, Jacobson EL, Ame J, Lin W;  
XX  
XX WPI; 2003-155895/15.  
XX DR N-PSDB; ABX14478.  
XX  
XX New nucleic acid molecule encoding a polypeptide with poly(ADP-ribose)

PT glycohydrolase activity, for preventing, treating, or ameliorating a  
PT disease condition, e.g. neoplastic disorder, myocardial infarction or  
PT vascular stroke.  
XX  
XX Claim 28; Fig 16; 86pp; English.  
XX  
XX The present invention relates to the isolation of poly(ADP-ribose)  
CC glycohydrolase (PARG) from several species, and the polynucleotide  
CC sequences encoding them. Methods for inhibiting PARG expression or  
CC overexpressing PARG are also disclosed. PARG is involved in the cellular  
CC response to DNA damage, and is associated with the body's response to  
CC neoplastic disorder inducing agents and oxidative stress. The  
CC polynucleotide sequences encoding PARG and PARG modulators are useful for  
CC preventing, treating, or ameliorating diseases such as neoplastic  
CC disorders, myocardial infarction, vascular stroke, neurodegenerative  
CC disorders (e.g. Alzheimer's disease, Parkinson's disease or Huntington's  
CC disease), inborn genetic errors, reperfusion following ischaemia, aging,  
CC and neurotoxicity. The polynucleotide sequences are also useful in gene  
CC therapy. The methods are useful for identifying an agent that modulates  
CC PARG activity, identifying a mutant PARG allele in an individual, or  
CC screening candidate molecules for PARG modulating activity. The present  
CC sequence represents human PARG enzyme  
XX  
XX Sequence 976 AA;  
SQ  
Query Match 100.0%; Score 167; DB 6; Length 976;  
Best Local Similarity 100.0%; Pred. No. 3.9e-15;  
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 LFTVLDHNECLIIITGTOYSEYTGVAETYSR 31  
DB 770 LFTVLDHNECLIIITGTOYSEYTGVAETYSR 800  
  
XX  
XX RESULT 14  
XX AAE25629  
XX ID AAE25629 standard; protein; 977 AA.  
XX AC AAE25629;  
XX DT 04-NOV-2002 (first entry)  
XX DE Bovine poly adenosine diphosphate-ribose glycohydrolase (bPARG).  
XX  
XX Bovine; poly adenosine diphosphate-ribose glycohydrolase; bPARG; enzyme;  
XX ADP-ribose; neoplastic disorder; genotoxic oxidative stress; neurotropic;  
XX cardiac disorder; neuronal disorder; reperfusion injury; neurotoxicity;  
XX Alzheimer's disease; Huntington's disease; Parkinson's disease; cardiatic;  
XX cytostatic; vasotropic; neuroprotective; anticonvulsant; gene therapy;  
XX antisense therapy.  
XX  
XX Bos taurus.  
XX OS  
XX US6395543-B1.  
XX PN  
XX 28-MAY-2002.  
XX PD  
XX  
XX 23-FEB-2000; 2000US-00511507.  
XX PF  
XX 01-MAY-1998; 98US-0083768P.  
XX PR  
XX 30-APR-1999; 99US-00302812.  
XX PR  
XX (KENT ) UNIV KENTUCKY RES FOUND.  
XX PA Jacobson MK, Jacobson EL, Ame J, Lin W;  
XX  
XX WPI; 2002-535641/57.  
XX DR N-PSDB; AAD42081.  
XX  
XX New nucleic acid molecule encoding bovine poly adenosine diphosphate-  
XX ribose glycohydrolase involved in cellular response to DNA damage,  
XX inhibition of which is useful for treating neoplastic disorders and  
XX neurodegenerative diseases.  
XX PT

XX Claim 3; Col 47-45; 77pp; English.

XX The invention relates to an isolated nucleic acid molecule which encodes

CC a bovine poly (adenosine diphosphate (ADP)-ribose) glycohydrolase (bPARG)

CC which catalyses release of ADP-ribose from an ADP ribose polymer. The

CC invention is useful as probes and primer molecules that can be used in

CC hybridisation assays and polymerase chain reaction (PCR) amplification.

CC The knowledge of the nucleotide sequence of the PARG gene permits the

CC preparation of antisense therapeutics containing sequences complementary

CC to the mRNA of PARG gene. The antisense therapeutics are useful to treat

CC neoplastic disorders and conditions caused by genotoxic oxidative stress

CC e.g., cardiac disorders, neuronal disorders, reperfusion injury,

CC neurotoxicity, Alzheimer's disease, Huntington's disease and Parkinson's

CC disease. The invention is useful in gene therapy and antisense therapy.

XX The present sequence is bPARG

XX Sequence 977 AA;

Query Match 100.0%; Score 167; DB 5; Length 977;

Best Local Similarity 100.0%; Pred. No. 3.9e-15;

Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LFTVLDHNECLITGTQYSEYTGVAETVR 31

DB 771 LFTVLDHNECLITGTQYSEYTGVAETVR 801

RESULT 15

AAU76020

ID AAU76020 standard; protein; 977 AA.

XX AC AAU76020;

XX 08-MAY-2002 (first entry)

XX Bovine poly (ADP-ribose) glycohydrolase (PARG) protein sequence.

XX Cow; poly (ADP-ribose) glycohydrolase; PARG; DNA repair; ADP-ribose;

KW adenosine diphosphate-ribose; neoplastic disorder; adenocarcinoma;

KW leukaemia; lymphoma; melanoma; myeloma; sarcoma; teratocarcinoma;

KW hyperplasia; hypertrophy; reperfusion; ischaemia; heart attack; stroke;

KW neurodegenerative disease; neurological disorder; Alzheimer's disease;

KW Huntington's disease; Parkinson's disease.

XX OS Bos taurus.

XX Key Location/Qualifiers

FT Region 601..617

FT /note= "Represents PARG oligopeptide #1"

FT Region 761..770

FT /note= "Represents PARG oligopeptide #2"

FT Region 771..801

FT /note= "Represents PARG oligopeptide #3"

FT Region 849..880

FT /note= "Represents PARG oligopeptide #4"

XX US6337202-B1.

XX 08-JAN-2002.

XX 23-FEB-2000; 2000US-00511477.

XX 01-MAY-1998; 98US-0083768P.

XX 30-APR-1999; 99US-00302812.

XX (KENT ) UNIV KENTUCKY RES FOUND.

XX Jacobson MK, Jacobson EL, Ame J, Lin W;

XX WPI; 2002-163240/21.

XX N-PSDB; ABK14931.

PT Novel isolated and purified poly (ADP-ribose) glycohydrolase protein

PT which catalyses release of ADP-ribose from ADP ribose polymer, useful for

PT treating neoplastic and neurological disorders, heart attack and stroke.

XX Claim 2; Col 47-52; 81pp; English.

XX The present invention relates to a new poly (ADP-ribose) glycohydrolase

CC (PARG) protein which catalyses release of ADP-ribose from an ADP

CC (adenosine diphosphate)-ribose polymer. The PARG molecule of the

CC invention is useful for generating antibodies and can be inhibited or

CC activated for diagnosing and treating neoplastic disorders such as

CC adenocarcinoma, leukaemia, lymphoma, melanoma, myeloma, sarcoma,

CC teratocarcinoma, hyperplasia and hypertrophy, reperfusion following

CC ischaemia, heart attack, stroke, neurodegenerative diseases, neurological

CC disorders including Alzheimer's, Huntington's and Parkinson's diseases,

CC and related conditions. PARG levels may be enhanced to suppress DNA

CC repair and increase the cell's susceptibility to chemotherapy drugs.

CC Antagonists of PARG are administered to treat or prevent neoplastic

CC disorders. The present amino acid sequence represents the bovine PARG

CC protein of the invention. This protein is one of several PARG proteins

CC (AAU76020-AAU76024) of the invention

XX Sequence 977 AA;

Query Match 100.0%; Score 167; DB 5; Length 977;

Best Local Similarity 100.0%; Pred. No. 3.9e-15;

Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LFTVLDHNECLITGTQYSEYTGVAETVR 31

DB 771 LFTVLDHNECLITGTQYSEYTGVAETVR 801

Search completed: May 26, 2004, 18:40:08

Job time : 4.55169 secs



US-09-973-451-2  
; Sequence 2, Application US/09973451  
; Patent No. US20020132328A1  
; GENERAL INFORMATION:  
; APPLICANT: JACOBSON, Myron K.  
; APPLICANT: JACOBSON, Elaine L.  
; APPLICANT: AM, Jean-Christophe  
; APPLICANT: LIN, Winston  
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE  
; TITLE OF INVENTION: (PARG) ENZYMES,  
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV  
; TITLE OF INVENTION: THEREWITH  
; FILE REFERENCE: NIAD 201  
; CURRENT APPLICATION NUMBER: US/09/973,451  
; CURRENT FILING DATE: 2001-10-09  
; PRIOR APPLICATION NUMBER: US/09/302,812  
; PRIOR FILING DATE: 1999-04-30  
; PRIOR APPLICATION NUMBER: 60/083,768  
; PRIOR FILING DATE: 1998-05-01  
; NUMBER OF SEQ ID NOS: 38  
; SEQ ID NO 2  
; LENGTH: 977  
; TYPE: PRT  
; ORGANISM: Bos taurus  
; FEATURE:  
US-09-973-451-2  
Query Match 100.0%; Score 167; DB 9; Length 977;  
Best Local Similarity 100.0%; Pred. No. 1.8e-16;  
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 LFTVLDHNECLIIITGTEQYSEYTGVAETYR 31  
DB 771 LFTVLDHNECLIIITGTEQYSEYTGVAETYR 801  
RESULT 5  
US-09-973-451-8  
; Sequence 8, Application US/09973451  
; Patent No. US20020132328A1  
; GENERAL INFORMATION:  
; APPLICANT: JACOBSON, Myron K.  
; APPLICANT: JACOBSON, Elaine L.  
; APPLICANT: AM, Jean-Christophe  
; APPLICANT: LIN, Winston  
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE  
; TITLE OF INVENTION: (PARG) ENZYMES,  
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV  
; TITLE OF INVENTION: THEREWITH  
; FILE REFERENCE: NIAD 201  
; CURRENT APPLICATION NUMBER: US/09/973,451  
; CURRENT FILING DATE: 2001-10-09  
; PRIOR APPLICATION NUMBER: US/09/302,812  
; PRIOR FILING DATE: 1999-04-30  
; PRIOR APPLICATION NUMBER: 60/083,768  
; PRIOR FILING DATE: 1998-05-01  
; NUMBER OF SEQ ID NOS: 38  
; SEQ ID NO 8  
; LENGTH: 768  
; TYPE: PRT  
; ORGANISM: Drosophila melanogaster  
; FEATURE:  
US-09-973-451-8  
Query Match 49.1%; Score 82; DB 9; Length 768;  
Best Local Similarity 53.3%; Pred. No. 0.0013;  
Matches 16; Conservative 5; Mismatches 9; Indels 0; Gaps 0;  
QY 1 LFTVLDHNECLIIITGTEQYSEYTGVAETY 30  
DB 400 LFTVLDHNECLIIITGTEQYSEYTGVAET 429

US-09-973-451-6  
; Sequence 6, Application US/09973451  
; Patent No. US20020132328A1  
; GENERAL INFORMATION:  
; APPLICANT: JACOBSON, Myron K.  
; APPLICANT: JACOBSON, Elaine L.  
; APPLICANT: AM, Jean-Christophe  
; APPLICANT: LIN, Winston  
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE  
; TITLE OF INVENTION: (PARG) ENZYMES,  
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV  
; TITLE OF INVENTION: THEREWITH  
; FILE REFERENCE: NIAD 201  
; CURRENT APPLICATION NUMBER: US/09/973,451  
; CURRENT FILING DATE: 2001-10-09  
; PRIOR APPLICATION NUMBER: US/09/302,812  
; PRIOR FILING DATE: 1999-04-30  
; PRIOR APPLICATION NUMBER: 60/083,768  
; PRIOR FILING DATE: 1998-05-01  
; NUMBER OF SEQ ID NOS: 38  
; SEQ ID NO 6  
; LENGTH: 968  
; TYPE: PRT  
; ORGANISM: Mus musculus  
; FEATURE:  
US-09-973-451-6  
Query Match 100.0%; Score 167; DB 9; Length 968;  
Best Local Similarity 100.0%; Pred. No. 1.7e-16;  
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 LFTVLDHNECLIIITGTEQYSEYTGVAETYR 31  
DB 762 LFTVLDHNECLIIITGTEQYSEYTGVAETYR 792  
RESULT 3  
US-09-973-451-4  
; Sequence 4, Application US/09973451  
; Patent No. US20020132328A1  
; GENERAL INFORMATION:  
; APPLICANT: JACOBSON, Myron K.  
; APPLICANT: JACOBSON, Elaine L.  
; APPLICANT: AM, Jean-Christophe  
; APPLICANT: LIN, Winston  
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE  
; TITLE OF INVENTION: (PARG) ENZYMES,  
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV  
; TITLE OF INVENTION: THEREWITH  
; FILE REFERENCE: NIAD 201  
; CURRENT APPLICATION NUMBER: US/09/973,451  
; CURRENT FILING DATE: 2001-10-09  
; PRIOR APPLICATION NUMBER: US/09/302,812  
; PRIOR FILING DATE: 1999-04-30  
; PRIOR APPLICATION NUMBER: 60/083,768  
; PRIOR FILING DATE: 1998-05-01  
; NUMBER OF SEQ ID NOS: 38  
; SEQ ID NO 4  
; LENGTH: 976  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
US-09-973-451-4  
Query Match 100.0%; Score 167; DB 9; Length 976;  
Best Local Similarity 100.0%; Pred. No. 1.8e-16;  
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 LFTVLDHNECLIIITGTEQYSEYTGVAETYR 31  
DB 770 LFTVLDHNECLIIITGTEQYSEYTGVAETYR 800

RESULT 8  
US-10-425-114-65345  
; Sequence 65345, Application US/10425114  
; Publication No. US20040034888A1  
; GENERAL INFORMATION:  
; APPLICANT: Liu, Jingdong



```

: FEATURE:
US-9-973-451-10
Query Match 31.7%; Score 53; DB 9; Length 726;
Best Local Similarity 35.5%; Pred. No. 32;
Matches 11; Conservative 4; Mismatches 16; Indels 0; Gaps 0;

QY 1 LFTVLVDHNECLIIITGTEQYSEYVTGYAETVR 31
DB 552 LLCEKMKQLEAISIVGAYVFSSYTGYGTLK 582

RESULT 12
US-10-282-122A-71975
; Sequence 71975, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangau
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 71975
; LENGTH: 307
; TYPE: PR1
; ORGANISM: Staphylococcus haemolyticus
US-10-282-122A-71975

Query Match 29.6%; Score 49.5; DB 12; Length 307;
Best Local Similarity 50.0%; Pred. No. 39;
Matches 13; Conservative 3; Mismatches 9; Indels 1; Gaps 1;

QY 1 LFTVLVDHNECLIIITGTEQYSEYTG 25
DB 118 LFTVKYPIAQGLIITGTGYPKYG 143

RESULT 13
US-10-425-114-45983

```

```
; Sequence 45983, Application US/10425114
; Publication No. US2004003488A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 45983
; LENGTH: 224
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: 700894080_FLI.pep
US-10-425-114-45983

Query Match      29.3%; Score 49; DB 12; Length 224;
Best Local Similarity 40.0%; Pred. No. 32;
Matches      8; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

QY      7 DHNECLIIITGTEQYSEYTYG 26
      |||:|:|:|:|:|:|:|:|
DB      146 DHDMCVPTGSEAWTRSLGY 165

RESULT 14
US-10-425-114-52216
; Sequence 52216, Application US/10425114
; Publication No. US2004003488A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 52216
; LENGTH: 329
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: 700788604_FLI.pep
US-10-425-114-52216

Query Match      29.3%; Score 49; DB 12; Length 329;
Best Local Similarity 40.0%; Pred. No. 51;
Matches      8; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

QY      7 DHNECLIIITGTEQYSEYTYG 26
      |||:|:|:|:|:|:|:|:|
DB      251 DHDMCVPTGSEAWTRSLGY 270

RESULT 15
US-10-389-566-1764
; Sequence 1764, Application US/10389566
; Publication No. US20040025202A1
; GENERAL INFORMATION:
; APPLICANT: Monsanto Technology, LLC
; APPLICANT: Laurie, Cathy C
```

```
; TITLE OF INVENTION: Nucleic Acid Molecules Associated with Oil in Plants
; FILE REFERENCE: 38-77(52900)D
; CURRENT APPLICATION NUMBER: US/10/389,566
; CURRENT FILING DATE: 2003-03-31
; PRIOR APPLICATION NUMBER: US 60/365,301
; PRIOR FILING DATE: 2002-03-15
; PRIOR APPLICATION NUMBER: US 60/391,786
; PRIOR FILING DATE: 2002-06-25
; PRIOR APPLICATION NUMBER: US 60/392,018
; PRIOR FILING DATE: 2002-06-26
; NUMBER OF SEQ ID NOS: 2459
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1764
; LENGTH: 427
; TYPE: PRT
; ORGANISM: Trichodesmium erythraeum
US-10-389-566-1764

Query Match      29.3%; Score 49; DB 16; Length 427;
Best Local Similarity 36.0%; Pred. No. 70;
Matches      9; Conservative 4; Mismatches 12; Indels 0; Gaps 0;

QY      5 VLDHNECLIIITGTEQYSEYTYG 29
      |||:|:|:|:|:|:|:|:|
DB      254 ILLEECALSRGARIYAEMVGYGT 278

Search completed: May 26, 2004, 19:19:00
Job time : 3.58646 secs
```

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 26, 2004, 17:20:49 ; Search time 0.496675 Seconds  
(without alignments)  
3249.964 Million cell updates/sec

Title: US-09-302-812-11  
Perfect score: 167  
Sequence: 1 LFTVELDNECLITGTQSEYTGVAETIR 31

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_42.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	52	31.1	510	1 CBP1 ORYSA	P37890 oryza sativ
2	50.5	30.2	541	1 YB36 METJA	Q58536 methanococc
3	50	29.9	364	1 VPAP ESVUJ	P52440 human herpe
4	50	29.9	499	1 CBP1 HORVU	P07519 hordeum vul
5	48.5	29.0	558	1 LEU1 CLOAB	Q97mc5 clostridium
6	48	28.7	250	1 YM25 ARCFU	O28058 archaeoglob
7	48	28.7	474	1 VP61 NPVOP	O10270 oryza pseu
8	48	28.7	521	1 NU2C SYNTP	P29801 synchococc
9	47.5	28.4	844	1 MCEL VACCC	P20979 vaccinia vi
10	47.5	28.4	844	1 MCEL VACCV	P04298 vaccinia vi
11	47.5	28.4	2186	1 YL52 CABEL	P34431 caenorhabdi
12	47	28.1	520	1 NU2C ANASP	Q8ymq0 anabaena sp
13	47	28.1	693	1 REGG PASMU	Q9cmb4 pasteurella
14	46.5	27.8	164	1 PHEA PORPU	P51368 porphyra pu
15	46.5	27.8	164	1 PHEA PORTE	O49843 porphyra te
16	46.5	27.8	164	1 PHEA PORYE	O20206 porphyra ye
17	46.5	27.8	164	1 PHEA RHOVL	Q02036 rhodella vi
18	46.5	27.8	401	1 ASSY STAAW	Q89xvc7 staphylococ
19	46.5	27.8	401	1 ASSY STAAW	Q42937 schizosacch
20	46.5	27.8	796	1 CORP SCHPO	P50851 homo sapien
21	46.5	27.8	2863	1 LRBA HUMAN	P37953 bacillus su
22	46	27.5	76	1 CSBA BACSU	Q00758 bacillus su
23	46	27.5	518	1 SP5B BACSU	Q9z6x5 chlamydia p
24	46	27.5	527	1 SYK CHLPN	Q09573 caenorhabdi
25	45.5	27.2	1020	1 YRD3 CABEL	P11045 bacillus su
26	45	26.9	168	1 DYL BACSU	P35003 haliotis ru
27	45	26.9	254	1 CTRL HALRU	P50147 gallus gall
28	45	26.9	354	1 GB12 CHICK	Q9chw7 lactococcus
29	45	26.9	415	1 GLYA LACLA	Q9r112 mus musculu
30	45	26.9	450	1 SORD MOUSE	P31723 penicillium
31	45	26.9	511	1 MAL2 PNCNI	Q94501 schizosacch
32	45	26.9	556	1 TCPA SCHPO	P05335 candida mal
33	45	26.9	708	1 CAO4 CANNA	

34	45	26.9	723	1 CAO4 CANTR	P11356 candida tro
35	45	26.9	797	1 AF32 HUMAN	Q9v4w6 homo sapien
36	45	26.9	1040	1 RIK1 SCHPO	Q10426 schizosacch
37	45	26.9	1390	1 INSR AEDAE	Q93105 aedes aegypt
38	44.5	26.6	598	1 FLIF PSRAE	Q51463 pseudomonas
39	44.5	26.6	742	1 ZW10 ARATH	O48626 arabidopsis
40	44.5	26.6	1000	1 COPP CABEL	Q20168 caenorhabdi
41	44	26.3	97	1 CH10 BUCTC	Q9f4e8 buchnera ap
42	44	26.3	137	1 COPI DICDI	P54706 dictyosteli
43	44	26.3	161	1 UREE PROMI	P17090 proteus mir
44	44	26.3	164	1 PHAI SYNFW	Q08086 synchococc
45	44	26.3	164	1 PHAI SYNFP	Q02179 synchococc

## ALIGNMENTS

RESULT 1					
ID	CBP1 ORYSA	STANDARD;	PRT;	510 AA.	
AC	P37890;				
DT	01-OCT-1994 (Rel. 30, Created)				
DT	01-OCT-1994 (Rel. 30, Last sequence update)				
DT	10-OCT-2003 (Rel. 42, Last annotation update)				
DE	Serine carboxypeptidase I precursor (EC 3.4.16.5) (Carboxypeptidase C)				
GN	CBP1.				
OS	Oryza sativa (Rice).				
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;				
OC	Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;				
OC	Ehrhartoideae; Oryzae; Oryza.				
OX	NCBI_TaxID=4530;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=cv. Yukihikari;				
RX	MEDLINE=94213891; PubMed=8161571;				
RA	Washio K., Ishikawa K.;				
RT	"Cloning and sequencing of the gene for type I carboxypeptidase in Rice."				
RL	Biochim. Biophys. Acta 1199:311-314(1994).				
CC	-!- CATALYTIC ACTIVITY: Release of a C-terminal amino acid with a broad specificity.				
CC	-!- PTM: Three disulfide bonds are present (Potential).				
CC	-!- SIMILARITY: Belongs to peptidase family S10.				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).				
CC	EMBL; D17596; BAA04510.1; -.				
DR	PIR; S43516; S43516.				
DR	HSSP; P08819; 1WHT.				
DR	MEROPS; S10.004; -.				
DR	Gramene; P37890; -.				
DR	InterPro; IPR001563; Peptidase_S10.				
DR	InterPro; IPR000379; Ser esters.				
DR	Pfam; PF00450; serine carboxpept; 1.				
DR	PRINTS; PR00724; CRBOXYPTASEC.				
DR	ProDom; PD001189; Serine carboxpept; 2.				
DR	PROSITE; PS00131; CARBOXYPEPT_SER_SER; 1.				
DR	PROSITE; PS00560; CARBOXYPEPT_SER_HIS; 1.				
DR	Hydrolase; Carboxypeptidase; Glycoprotein; Zymogen; Signal.				
KW	SIGNAL 1 25				
FT	POTENTIAL.				
FT	PROPEP 26 36				
FT	CHAIN 37 510				
FT	ACT SITE 194 194				
FT	ACT SITE 434 434				
FT	ACT SITE 487 487				
FT	CARBOHYD 154 154				
FT	N-LINKED (GLCNAC. . .) (POTENTIAL).				

QY  
1 LFTFVLVDHNE-----CLITGTEQY-----SEYTGVAE 28

RK  
292 MEKTEIFENDNDFKNMKVIYDCTVIEGTETELYMNIKRGEVKPYRE 33

"The A- and B-chains of carboxypeptidase I from germinated barley originate from a single precursor polypeptide.";  
J. Biol. Chem. 263:11106-11110(1988).  
[3]  
SEQUENCE OF 31-296 AND 322-499.  
Scerensen S.B., Breddam K., Svendsen I.;  
RA "Primary structure of carboxypeptidase I from malted barley."; Carlsberg Res. Commun. 51:475-485(1986).  
RL  
CC -|- FUNCTION: May be involved in the degradation of small peptides (2-5 residues) or in the degradation of storage proteins in the embryo.  
CC  
CC -|- CATALYTIC ACTIVITY: Release of a C-terminal amino acid with a broad specificity.  
CC  
CC -|- SUBUNIT: Carboxypeptidase I is a dimer, where each monomer is composed of two chains linked by disulfide bonds.  
CC  
CC -|- SUBCELLULAR LOCATION: Secreted into the endoplasmic reticulum.  
CC  
CC -|- DEVELOPMENTAL STAGE: After one day of germination, mainly found in the scutellum of the developing grain; barely detectable after four days, and absent from the mature grain. A lower level of expression is seen in the aleurone both during development and germination.  
CC  
CC -|- PTM: Three disulfide bonds are present.  
CC  
CC -|- PTM: THE LINKER PEPTIDE IS ENDOPEPTIDOLYTIALLY EXCISED DURING ENZYME MATURATION.  
CC  
CC -|- SIMILARITY: Belongs to peptidase family S10.  
-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announcement/ or send an email to license@isb-sib.ch).  
-----  
EMBL; Y03603; CAA70816.1; ;  
DR EMBL; J03897; AAA32940.1; ;  
DR PIR; T05367; CPBHS.  
DR HSP; P08819; 1WHT.  
DR MEROPS; S10.004; ;  
DR InterPro; IPR001563; Peptidase\_S10.  
DR InterPro; IPR000379; Ser\_estr.  
DR Pfam; PF00450; serine\_carbpept; 1.  
DR PRINTS; PR00724; CRBOXYPATSEC.  
DR ProDom; PD001189; Serine carbpept; 2.  
DR PROSITE; PS00131; CARBOXYPEPT\_SER\_HIS; 1.  
DR PROSITE; PS00560; CARBOXYPEPT\_SER\_HIS; 1.  
KW Hydrolase; Carboxypeptidase; Glycoprotein; Zymogen; Signal.  
FT SIGNAL 1 30  
FT CHAIN 31 296 SERINE CARBOXYPEPTIDASE I, CHAIN A.  
FT PROPEP 297 351 LINKER PEPTIDE.  
FT CHAIN 352 499 SERINE CARBOXYPEPTIDASE I, CHAIN B.  
FT ACT\_SITE 188 188 BY SIMILARITY.  
FT ACT\_SITE 423 423 BY SIMILARITY.  
FT ACT\_SITE 476 476 BY SIMILARITY.  
FT CARBOHYD 148 148 N-LINKED (GLCNAC. . .).  
FT CARBOHYD 262 262 N-LINKED (GLCNAC. . .).  
FT CARBOHYD 407 407 N-LINKED (GLCNAC. . .).  
FT SITE 497 499 MICROBODY TARGETING SIGNAL (POTENTIAL).  
FT CONFLICT 102 102 H -> P (IN REF. 3).  
SQ SEQUENCE 499 AA; 54096 MW; 9C6674B14D9DB9BF CRC64;

Query Match 29.9%; Score 50; DB 1; Length 499;  
Best Local Similarity 40.0%; Pred. No. 15;  
Matches 8; Conservative 6; Mismatches 6; Indels 0; Gaps 0;

Qy 7 DHNECLITGTQSYETCY 26  
||: ||: ||: ||: ||: ||:  
Db 421 DHDMCVFPTGSEAWTKSLGY 440

RESULT 5  
LEU1\_CLOAB STANDARD. OPT. 558 AA

AC Q97MC5;  
DT 28-FEB-2003 (Rel. 41, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DL J. Biol. Chem. 263:11106-11110(1988).  
DE 2-isopropylmalate synthase (EC 2.3.3.13) (Alpha-isopropylmalate synthase) (Alpha-IPM synthetase).  
DN LEUA OR CAC0273.  
OS Clostridium acetobutylicum.  
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae; Clostridium.  
OX NCBI\_TaxId=1488;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX STRAIN=ATCC 824 / DSM 792 / VKM B-1787;  
RA NOELLING J., Bretton G., Omelchenko M.V., Makarova K.S., Zeng Q., Gibson R., Lee H.M., Dubois J., Qiu D., Hitti J., Wolf Y.I., Tatunov R.L., Koonin E.V., Doucette-Stamm L., Soucaille P., Daly M.J., Bennett G.N., Koonin E.V., Smith D.R.;  
RA "Genome sequence and comparative analysis of the solvent-producing bacterium Clostridium acetobutylicum";  
RT Bacteriol. 183:4823-4838(2001).  
RL J. Bacteriol. 183:4823-4838(2001).  
CC -|- FUNCTION: Catalyzes the condensation of the acetyl group of acetyl-CoA with 3-methyl-2-oxobutanolate (2-oxoisovalerate) to form 3-carboxy-3-hydroxy-4-methylpentanoate (2-isopropylmalate).  
CC -|- CATALYTIC ACTIVITY: Acetyl-CoA + 3-methyl-2-oxobutanolate + H(2)O = 2-hydroxy-2-isopropylsuccinate + CoA.  
CC -|- PATHWAY: Leucine biosynthesis; first step.  
CC -|- SUBUNIT: Homotetramer (By similarity).  
CC -|- SIMILARITY: Belongs to the alpha-IPM synthetase / homocitrate synthase family. LeuA 2 subfamily.  
-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announcement/ or send an email to license@isb-sib.ch).  
-----  
EMBL; AE007540; AAK78254.1; ;  
DR PIR; C96933; C96933.  
DR HAMAP; MF\_00572; -1.  
DR InterPro; IPR002034; ALPM/Hcit\_synth.  
DR InterPro; IPR000891; HMGL-like.  
DR InterPro; IPR005669; LeuA yeast.  
DR Pfam; PF00682; HMGL-like; 1.  
DR TIGRfam; TIGR00970; leuA yeast; 1.  
DR PROSITE; PS00815; ALPM HOMOCIT\_SYNTH\_1; 1.  
DR PROSITE; PS00816; ALPM HOMOCIT\_SYNTH\_2; FALSE NEG.  
DR Leucine biosynthesis; Transferrase; Complete proteome.  
SQ SEQUENCE 558 AA; 63443 MW; 3573DD2616A09B3F CRC64;

Query Match 29.0%; Score 48.5; DB 1; Length 558;  
Best Local Similarity 37.5%; Pred. No. 29;  
Matches 12; Conservative 7; Mismatches 10; Indels 3; Gaps 1;

Qy 1 LFTEVDHNECLITGTE---QYSEVYGTAET 29  
:||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:  
Db 136 VFKKDKDEVKSLAIKGAQMVKYSEYTEYSES 167

RESULT 6  
YM25\_ARCFU STANDARD; PRT; 250 AA.  
ID YM25\_ARCFU  
AC O28058;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DL 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Hypothetical protein AF2225.  
DN AF2225.  
OS Archaeoglobus fulgidus.  
OC Archaea; Euryarchaeota; Archaeoglobi; Archaeoglobales;

```

Archaeoglobaceae; Archaeoglobus.
[1]
NCBI_TaxID=2234;
SEQUENCE FROM N.A.
STRAIN=VC-16 / DSM 4304 / ATCC 49558;
MEDLINE=98049343; PubMed=9388745;
Klenk H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,
Ketchum K.A., Dodson R.J., Gwinn M., Hickey E.K., Peterson J.D.,
Richardson D.L., Kurlavsky A.R., Graham D.E., Kyrpides N.C.,
Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,
Kirkness E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,
Peterson S., Reich C.I., McNeil L.K., Badger J.H., Glodek A., Zhou L.,
Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Utterback T.,
Cotton M.D., Spriggs T., Artiach P., Kaine B.P., Sykes S.M.,
Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,
Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,
Venter J.C.;
"The complete genome sequence of the hyperthermophilic, sulphate-
reducing archaeon Archaeoglobus fulgidus.";
Nature 390:364-370(1997).
-!- SIMILARITY: Belongs to the FAH family.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; AE000952; AAB89031.1; -.
PIR; A69528; A69528.
TIGR; AF2225; -.
InterPro; IPR002529; FAA_hydrolase.
Pfam; PF01557; FAA_hydrolase; 1.
Hypothetical protein; Complete proteome.
SEQUENCE 250 AA; 27370 MW; AD0E196CA19C6EC0 CRC64;
-----
Query Match 28.7%; Score 48; DB 1; Length 250;
Best Local Similarity 34.8%; Pred.No. 14;
Matches 8; Conservative 6; Mismatches 9; Indels 0; Gaps 0;
-----
QY 3 TEVLHDNECLIIITGEQSYEYTG 25
| | : : : : :
Db 76 TAVIGHDDCIILPQISQVDYEG 98
| | : : : : :
RESULT 7
VP61 NPVOP STANDARD; PRY 474 AA.
ID ID VP61 NPVOP STANDARD; PRY 474 AA.
AC 010270;
AT 01-NOV-1997 (Rel. 35, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE 61 kDa protein homolog.
OS Orygia pseudotsugata multicapsid polyhedrosis virus (OpMPV).
OC Viruses; dsDNA viruses, no RNA stage; Baculoviridae;
OC Nucleopolyhedrovirus.
OX NCBI_TaxID=164623;
ON [1]
RN
RP SEQUENCE FROM N.A.
RX MEDLINE=97271300; PubMed=9126251;
RA Ahrens C.H., Russell R.R., Funk C.J., Evans J., Harwood S.,
RA Rohrmann G.F.;
RT "The sequence of the Orygia pseudotsugata multinucleocapsid nuclear
polyhedrosis virus genome.";
RL Virology 229:381-393(1997).
[2]
RN
RP REVISIONS.
RA Rohrmann G.F.;
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: Probably plays an important role in the persistence and
survival of the virus. May be a structural component in the

```

```

DR PIR: S19921; DNYC27.
DR HAMAP; MF_00445; 1.
DR InterPro; IPR003916; NADhub_oxred5.
DR InterPro; IPR001750; Oxidored_q1.
DR Pfam; PF00361; oxidored_q1; 1.
DR PRINTS; PR01434; NADHDGNAS5.
DR Oxidoreductase; NAD; NADP; Quinone; Plastoquinone; Transmembrane.
FT TRANSMEM 14 34 POTENTIAL.
FT TRANSMEM 42 62 POTENTIAL.
FT TRANSMEM 79 99 POTENTIAL.
FT TRANSMEM 109 129 POTENTIAL.
FT TRANSMEM 132 152 POTENTIAL.
FT TRANSMEM 167 187 POTENTIAL.
FT TRANSMEM 207 227 POTENTIAL.
FT TRANSMEM 241 261 POTENTIAL.
FT TRANSMEM 275 295 POTENTIAL.
FT TRANSMEM 303 323 POTENTIAL.
FT TRANSMEM 331 351 POTENTIAL.
FT TRANSMEM 375 395 POTENTIAL.
FT TRANSMEM 397 417 POTENTIAL.
FT TRANSMEM 463 483 POTENTIAL.
FT TRANSMEM 495 515 POTENTIAL.
SQ SEQUENCE 521 AA; 55069 MW; 04FB335A066F84CC CRC64;

Query Match 28.7%; Score 48; DB 1; Length 521;
Best Local Similarity 52.6%; Pred. No. 31;
Matches 10; Conservative 3; Mismatches 2; Indels 4; Gaps 1;

QY 11 CLUI-----TGTEQYSEYTG 25
   ||| ||| ||| ||| |||
Db 347 CVILFSLRTGTDQISEVAG 365

RESULT 9
MCEL VACCV STANDARD; PRT; 844 AA.
AC P20979;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE mRNA capping enzyme large subunit [includes: polynucleotide 5'-
DE triphosphatase (EC 3.1.3.33) (mRNA 5'-triphosphatase) (TPase); mRNA
DE guanylyltransferase (EC 2.7.7.50) (GTP--RNA guanylyltransferase)
DE (Gtase)].
GN DIR.
OS Vaccinia virus (strain Copenhagen).
OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
OC Orthopoxvirus.
OX NCBI_TaxID=10249;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91021027; PubMed=2219722;
RA Goebel S.-J., Johnson G.P., Perkus M.E., Davis S.W., Winslow J.P.,
RA Paolletti E.;
RT "The complete DNA sequence of vaccinia virus.";
RL Virology 179:247-266(1990).
RN [2]
RP COMPLETE GENOME.
RA Goebel S.-J., Johnson G.P., Perkus M.E., Davis S.W., Winslow J.P.,
RA Paolletti E.;
RT "Appendix to 'The complete DNA sequence of vaccinia virus'.";
RL Virology 179:517-563(1990).
CC -|- FUNCTION: CATALYZES THE FIRST TWO REACTIONS IN THE MRNA CAP
CC FORMATION PATHWAY.
CC -|- CATALYTIC ACTIVITY: 5'-phosphopolynucleotide + H(2)O =
CC polynucleotide + phosphate.
CC -|- CATALYTIC ACTIVITY: GTP + (5')PP-pur-mRNA = diphosphate +
CC G(5')PPP-pur-mRNA.
CC -|- SUBUNIT: Heterodimer of a large and a small subunit.
CC -|- SIMILARITY: Belongs to the viral Gtase family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC DR EMBL; M15058; AAA48253.1; -.
CC PIR; A03872; QQVZ1.

```

```

CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC DR EMBL; M35027; AAA48095.1; -.
CC PIR; I42514; QOVZVC.
CC InterPro; IPR004971; Pox_MCEL.
CC Pfam; PF03291; Pox_MCEL; 1.
CC Transferase; Nucleotidyltransferase; Hydrolase; mRNA processing;
CC mRNA capping; Multifunctional enzyme.
FT ACT SITE 260 260 GUANYLYLATION SITE (POTENTIAL).
SQ SEQUENCE 844 AA; 96732 MW; 9982FDC819867277 CRC64;

Query Match 28.4%; Score 47.5; DB 1; Length 844;
Best Local Similarity 35.3%; Pred. No. 62;
Matches 12; Conservative 5; Mismatches 14; Indels 3; Gaps 1;

QY 1 LFEVLHNECLIIITGTEQ---YSEYTGAYETVR 31
   ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 109 LVTEPLHKECLRLSTEEHIFLDYKYGSSIR 142

RESULT 10
MCEL VACCV STANDARD; PRT; 844 AA.
AC P04298;
DT 20-MAR-1987 (Rel. 04, Created)
DT 20-MAR-1987 (Rel. 04, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE mRNA capping enzyme large subunit [includes: Polynucleotide 5'-
DE triphosphatase (EC 3.1.3.33) (mRNA 5'-triphosphatase) (TPase); mRNA
DE guanylyltransferase (EC 2.7.7.50) (GTP--RNA guanylyltransferase)
DE (Gtase)].
GN DIR.
OS Vaccinia virus (strain WR).
OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
OC Orthopoxvirus.
OX NCBI_TaxID=10254;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=86291159; PubMed=3739227;
RA Niles E.G., Condit R.C., Caro P., Davidson K., Matusick L., Seto J.;
RA "Nucleotide sequence and genetic map of the 16-kb vaccinia virus
RA HindIII D fragment.";
RL Virology 153:96-112(1986).
RN [2]
RP INTERACTION WITH SMALL SUBUNIT, AND SEQUENCE OF 1-20 AND 498-517.
RX MEDLINE=90272646; PubMed=2161527;
RA Guo P., Moss B.;
RT "Interaction and mutual stabilization of the two subunits of vaccinia
RT virus mRNA capping enzyme coexpressed in Escherichia coli.";
RL Proc. Natl. Acad. Sci. U.S.A. 87:4023-4027(1990).
CC -|- FUNCTION: CATALYZES THE FIRST TWO REACTIONS IN THE MRNA CAP
CC FORMATION PATHWAY.
CC -|- CATALYTIC ACTIVITY: 5'-phosphopolynucleotide + H(2)O =
CC polynucleotide + phosphate.
CC -|- CATALYTIC ACTIVITY: GTP + (5')PP-pur-mRNA = diphosphate +
CC G(5')PPP-pur-mRNA.
CC -|- SUBUNIT: Heterodimer of a large and a small subunit.
CC -|- SIMILARITY: Belongs to the viral Gtase family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC DR EMBL; M15058; AAA48253.1; -.
CC PIR; A03872; QQVZ1.

```

DR



DR EMBL; AP003597; BAB76582.1; -.  
DR PIR; AC2416; AC2416.  
DR HAMAP; MF 00445; -. 1.  
DR InterPro; IPR003916; NADHub oxford.  
DR InterPro; IPR001750; Oxidored\_q1.  
DR Pfam; PF00361; Oxidored\_q1; 1.  
DR PRINTS; PR01434; NADHDHGNASE5.  
KW Oxidoreductase; NAD; NADP; Quinone; Plastoquinone; Transmembrane;  
FT Complete proteome. 36  
FT TRANSMEM 15  
FT TRANSMEM 43  
FT TRANSMEM 65  
FT TRANSMEM 80  
FT TRANSMEM 109  
FT TRANSMEM 128  
FT TRANSMEM 132  
FT TRANSMEM 151  
FT TRANSMEM 164  
FT TRANSMEM 186  
FT TRANSMEM 201  
FT TRANSMEM 223  
FT TRANSMEM 243  
FT TRANSMEM 280  
FT TRANSMEM 299  
FT TRANSMEM 306  
FT TRANSMEM 333  
FT TRANSMEM 335  
FT TRANSMEM 375  
FT TRANSMEM 407  
FT TRANSMEM 429  
FT TRANSMEM 466  
FT TRANSMEM 488  
SQ SEQUENCE 520 AA; 55464 MW; AD61672D5548C5 CRC64;  
Query Match 28.1%; Score 47; DB 1; Length 520;  
Best Local Similarity 47.4%; Pred. No. 44;  
Matches 9; Conservative 5; Mismatches 1; Indels 4; Gaps 1;  
QY 11 CLII-----TGTEQYSYTG 25  
DB 350 CHILSLRTGTQIAEYSG 368

RESULT 13  
RECG\_PASMU STANDARD; PRT; 693 AA.  
AC Q9CMB4;  
DC 28-FEB-2003 (Rel. 41, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Arp-dependent DNA helicase recG (EC 3.6.1.-).  
GN RecG OR PM0919.  
OS Pasteurella multocida.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;  
OC Pasteurellaceae; Pasteurella.  
OX NCBI\_TaxID=747;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Pm70;  
RC MEDLINE=21145866; PubMed=11248100;  
RA May B.J., Zhang Q., Li L.L., Paustian M.L., Whittam T.S., Kapur V.;  
RT "Complete genomic sequence of Pasteurella multocida Pm70.";  
RT Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465(2001).  
CC -!- FUNCTION: Critical role in recombination and DNA repair. Help  
process Holliday junction intermediates to mature products by  
catalyzing branch migration. Has a DNA unwinding activity  
characteristic of a DNA helicase with a 3' to 5' polarity. RecG  
unwind branched duplex DNA (V-DNA) (By similarity).  
CC -!- SIMILARITY: Belongs to the helicase family. RecG subfamily.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC EMBL; AE006131; AAK03003.1; -.  
CC InterPro; IPR001410; DEAD.  
CC InterPro; IPR001650; Helicase\_C.  
CC InterPro; IPR008994; Nucleic\_acid\_OR.

DR InterPro; IPR004609; RecG.  
DR InterPro; IPR004365; trna\_anti.  
DR Pfam; PF00270; DEAD; 1.  
DR Pfam; PF00271; Helicase\_C; 1.  
DR Pfam; PF01336; trna\_anti; 1.  
DR SMART; SM00487; DEXDc; 1.  
DR SMART; SM00490; HELICc; 1.  
DR TIGRFAMs; TIGR00643; recG; 1.  
KW Hydrolyase; Helicase; DNA repair; ATP-binding; DNA recombination;  
KW DNA-binding; Complete proteome.  
FT NP BIND 296 303 ATP (POTENTIAL).  
FT SITE 397 400 DEQH BOX.  
SQ SEQUENCE 693 AA; 78025 MW; 5AB96E4B5BA54DB3 CRC64;  
Query Match 28.1%; Score 47; DB 1; Length 693;  
Best Local Similarity 40.0%; Pred. No. 59;  
Matches 10; Conservative 6; Mismatches 7; Indels 2; Gaps 1;  
QY 1 LFTVELDNECI--IITGTEQYSEY 23  
DB 225 IFELLAHNLAMQKVRIGTQFSAY 249  
RESULT 14  
PHEA\_PORPU STANDARD; PRT; 164 AA.  
ID PHEA\_PORPU  
AC P51368;  
DT 01-OCT-1996 (Rel. 34, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE R-phycocerythrin alpha chain.  
GN CPEA.  
OS Porphyra purpurea.  
OG Chloroplast.  
OC Eukaryota; Rhodophyta; Bangiophyceae; Bangiales; Bangiaceae; Porphyra.  
OX NCBI\_TaxID=2787;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Avonport;  
RA Reith M.E., Munholland J.;  
RT "Complete nucleotide sequence of the Porphyra purpurea chloroplast  
genome.";  
RL Plant Mol. Biol. Rep. 13:333-335(1995).  
CC -!- FUNCTION: Light-harvesting photosynthetic bile pigment-protein  
from the phycobiliprotein complex.  
CC -!- SUBUNIT: Heterodimer of an alpha and a beta chain.  
CC -!- SUBCELLULAR LOCATION: Periphery of the rods of the phycobilisome.  
CC -!- PTM: Contains one covalently linked bilin chromophore.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC EMBL; U38804; AAC08254.1; -.  
CC PIR; S73289; S73289.  
CC HSSP; O36005; 1B8D.  
DR InterPro; IPR001659; Phycobilisome.  
DR Pfam; PF00502; Phycobilisome; 1.  
DR ProDom; PD000340; Phycobilisome; 1.  
KW Phycobilisome; Electron transport; Photosynthesis; Bile pigment;  
KW Chloroplast. 82 82 PHYCOCYTHROBILIN CHROMOPHORE.  
FT BINDING 139 139 PHYCOCYTHROBILIN CHROMOPHORE.  
SQ SEQUENCE 164 AA; 17698 MW; 823E10F912074C5D CRC64;  
Query Match 27.8%; Score 46.5; DB 1; Length 164;  
Best Local Similarity 36.4%; Pred. No. 15;  
Matches 12; Conservative 4; Mismatches 10; Indels 7; Gaps 2;

Search completed: May 26, 2004, 18:41:17  
Job time : 2.49667 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 26, 2004, 18:18:25 ; Search time 2.474 Seconds

(without alignments)  
3953.539 Million cell updates/sec

Title: US-09-302-812-11

Perfect score: 167

Sequence: 1 LFTVLDHNECLITGTEQSYTGYAETFR 31

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: sp\_archaea.\*
- 2: sp\_bacteria.\*
- 3: sp\_fungi.\*
- 4: sp\_human.\*
- 5: sp\_invertebrate.\*
- 6: sp\_mammal.\*
- 7: sp\_mhc.\*
- 8: sp\_organelle.\*
- 9: sp\_phase.\*
- 10: sp\_plant.\*
- 11: sp\_rodent.\*
- 12: sp\_virus.\*
- 13: sp\_vertebrate.\*
- 14: sp\_unclassified.\*
- 15: sp\_virus.\*
- 16: sp\_bacteriap.\*
- 17: sp\_archaeap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	167	100.0	920	11 Q8CB72	Q8cb72 mus musculu
2	167	100.0	961	11 Q80YQ6	Q80yq6 mus musculu
3	167	100.0	968	11 Q88622	Q88622 mus musculu
4	167	100.0	972	11 Q90YM2	Q90ym2 rattus norv
5	167	100.0	976	4 Q9Y4W7	Q9y4w7 homo sapien
6	167	100.0	976	4 Q86W56	Q86w56 homo sapien
7	167	100.0	976	4 Q7Z742	Q7z742 homo sapien
8	167	100.0	977	6 Q02776	Q02776 bos taurus
9	82	49.1	723	5 Q96QN8	Q96qn8 drosophila
10	82	49.1	768	5 Q46043	Q46043 drosophila
11	79	47.3	548	10 Q9SKB3	Q9skb3 arabidopsis
12	69	41.3	522	10 Q8VYAL	Q8vyal arabidopsis
13	57	34.1	368	5 Q86G14	Q86g14 toxoplasma
14	55	32.9	485	5 Q9N5L4	Q9n5l4 caenorhabdi
15	53.5	32.0	437	17 Q8TR71	Q8tr71 methanosarc
16	53.5	32.0	941	16 Q8DFP9	Q8dfp9 vibrio vuln

17	53	31.7	764	5 Q19637	Q19637 caenorhabdi
18	53	31.7	781	5 Q867X0	Q867x0 caenorhabdi
19	52	31.1	464	2 Q9A165	Q9a165 erwinia rha
20	51	30.5	132	16 Q32092	Q32092 bacillus su
21	51	30.5	238	11 Q9D410	Q9d410 mus musculu
22	50.5	30.2	332	5 Q86GZ6	Q86gz6 rhipicephal
23	50.5	29.9	166	16 Q8ZCR0	Q8zce0 yersinia pe
24	50	29.9	364	10 Q9SKB4	Q9skb4 arabidopsis
25	50	29.9	456	10 Q9SV78	Q9sv78 arabidopsis
26	50	29.9	497	10 Q8L7B2	Q8l7b2 arabidopsis
27	50	29.9	1759	5 Q9SPL4	Q9spl4 trypanosoma
28	49	29.3	447	16 Q8DTC3	Q8dtc3 streptococc
29	49	29.3	504	10 Q9LSV8	Q9lsv8 arabidopsis
30	49	29.3	618	2 Q9F5E5	Q9f5e5 agrobacteri
31	49	29.3	2523	5 Q9VTP0	Q9vtp0 drosophila
32	48	28.7	364	2 Q8GCY0	Q8gcy0 bartonella
33	48	28.7	372	2 Q56044	Q56044 streptococc
34	48	28.7	384	2 Q8KIC9	Q8kic9 streptococc
35	48	28.7	392	2 Q9LC10	Q9lc10 streptomyce
36	48	28.7	413	16 Q8RGX4	Q8rgx4 fusbacteri
37	48	28.7	427	2 Q9FAR5	Q9far5 streptomyce
38	48	28.7	515	16 Q8DM56	Q8dmr5 synecococc
39	48	28.7	659	16 Q816J3	Q816j3 bacillus ce
40	48	28.7	1114	12 P89460	P89460 herpes simp
41	48	28.7	1925	5 Q812D1	Q812d1 plasmodium
42	47.5	28.4	276	16 Q8F4H5	Q8f4h5 leptospira
43	47.5	28.4	441	2 Q84FV7	Q84fv7 methylobact
44	47.5	28.4	843	12 Q8JLC5	Q8jlc5 ectromellia
45	47.5	28.4	844	12 Q57209	Q57209 vaccinia vi

#### ALIGNMENTS

#### RESULT 1

Q8CB72 ID Q8CB72 PRELIMINARY; PRT; 920 AA.  
AC Q8CB72;  
DT 01-MAR-2003 (TREMBLrel. 23, Created)  
DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)  
DT 01-JUN-2003 (TREMBLrel. 24, Last annotation update)  
DE Poly.  
GN PARG.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6J; TISSUE=Bone;  
RX MEDLINE=22354683; PubMed=12466851;  
RA The PANTOM Consortium,  
RA the RIKEN Genome Exploration Research Group Phase I & II Team;  
RT "Analysis of the mouse transcriptome based on functional annotation of  
RT 60,770 full-length cDNAs.";  
RL Nature 420:563-573(2002).  
DR EMBL; AK036656; BAC29519.1; -.  
DR MGD; MGI:1347094; Parg.  
DR InterPro; IPR007724; PARG.  
DR Pfam; PF05028; PARG; 1.  
SQ SEQUENCE 920 AA; 103799 MW; 379E779E121B09A1 CRC64;

Query Match 100.0%; Score 167; DB 11; Length 920;  
Best Local Similarity 100.0%; Pred. No. 2.5e-15;  
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LFTVLDHNECLITGTEQSYTGYAETFR 31

|||||  
763 LFTVLDHNECLITGTEQSYTGYAETFR 793

#### RESULT 2

Q80YQ6

RESULT 4	
Q9QYM2	
ID Q9QYM2	PRELIMINARY;
AC Q9QYM2;	PRT; 972 AA.

QY 1 LFTVELDHNELIITGTQYSEYTGAAETVR 31  
|||||  
770 LFTVELDHNELIITGTQYSEYTGAAETVR 800  
Dh

## RESULT 6

Q86W56  
ID Q86W56 PRELIMINARY; PRT; 976 AA.  
AC Q86W56;  
DT 01-JUN-2003 (TrEMBLrel. 24, Created)  
DE 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)  
DE 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Similar to poly (ADP-ribose) glycohydrolase.  
OS Homo sapiens (Human)  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC Tissue=Skin;  
RA Strausberg R.;  
RL Submitted (APR-2003) to the EMBL/GenBank/DBJ databases.  
DR EMBL; BC050560; AAH50560.1; -;  
DR GO; GO:0016787; F:hydrolase activity; IEA.  
DR InterPro; IPR007724; PARG.  
DR Pfam; PF05028; PARG; 1.  
KW Hydrolase.  
SQ SEQUENCE 976 AA; 111109 MW; D6646353C6D0180E CRC64;  
  
Query Match 100.0%; Score 167; DB 4; Length 976;  
Best Local Similarity 100.0%; Pred. No. 2.7e-15;  
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 LFTVLDHNECLITGTQYSEYTGATYR 31  
DB 770 LFTVLDHNECLITGTQYSEYTGATYR 800  
|||||  
[1]  
RP SEQUENCE FROM N.A.  
RC Tissue=Testis;  
RA Strausberg R.;  
RL Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.  
  
RESULT 7  
Q72742  
ID Q72742 PRELIMINARY; PRT; 976 AA.  
AC Q72742;  
DT 01-OCT-2003 (TrEMBLrel. 25, Created)  
DE 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)  
DE 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Hypothetical protein.  
OS Homo sapiens (Human)  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC Tissue=Testis;  
RA Strausberg R.L.; Feingold E.A.; Grouse L.H.; Derge J.G.;  
RA Klausner R.D.; Collins F.S.; Wagner L.; Shenmen C.M.; Schuler G.D.;  
RA Altschul S.F.; Zeeberg B.; Buetow K.H.; Schaefer C.F.; Bhat N.K.;  
RA Hopkins R.F.; Jordan H.; Moore T.; Max S.I.; Wang J.; Hsieh F.;  
RA Diatchenko L.; Marusina K.; Farmer A.A.; Rubin G.M.; Hong L.;  
RA Stapleton M.; Soares M.B.; Bonaldo M.F.; Casavant T.L.; Scheetz T.E.;  
RA Brownstein M.J.; Usdin T.B.; Toshiyuki S.; Carninci P.; Prange C.;  
RA Raha S.S.; Loquellano N.J.; Peters G.J.; Abramson R.D.; Mullaly S.J.;  
RA Bosak S.A.; McSwain P.A.; McKernan K.J.; Malek J.A.; Gunaratne P.H.;  
RA Richards S.; Worley K.C.; Hale S.; Garcia A.M.; Gay L.J.; Hulyk S.W.;  
RA Villalon D.K.; Muzny D.M.; Sodergren E.J.; Lu X.; Gibbs R.A.;  
RA Fahey J.; Helton E.; Ketterman M.; Madan A.; Rodrigues S.; Sanchez A.;  
RA Whiting M.; Madan A.; Young A.C.; Shevchenko Y.; Bouffard G.G.;  
RA Blakeley R.W.; Touchman J.W.; Green E.D.; Dickson M.C.;  
RA Rodriguez A.C.; Grimwood J.; Schmutz J.; Myers R.M.; Butterfield Y.S.;  
RA Krzywinski M.I.; Skalka U.; Smailus D.E.; Schnerch A.; Schein J.E.;  
RA Jones S.J.; Marra M.A.;  
RA "Generation and initial analysis of more than 15,000 full-length human  
RT and mouse cDNA sequences";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:16999-16993 (2002).  
[2]  
RP SEQUENCE FROM N.A.  
RC Tissue=Testis;  
RA Strausberg R.;  
RL Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.

DR EMBL; BC052966; AAH52966.1; -;  
KW Hypothetical protein.  
SQ SEQUENCE 976 AA; 111090 MW; 47C8EE1826C4A74A CRC64;  
  
Query Match 100.0%; Score 167; DB 4; Length 976;  
Best Local Similarity 100.0%; Pred. No. 2.7e-15;  
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 LFTVLDHNECLITGTQYSEYTGATYR 31  
DB 770 LFTVLDHNECLITGTQYSEYTGATYR 800  
|||||  
[1]  
RP SEQUENCE FROM N.A.  
RC Tissue=Testis;  
RA Strausberg R.;  
RL Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.  
  
RESULT 8  
O02776  
ID O02776 PRELIMINARY; PRT; 977 AA.  
AC O02776;  
DT 01-JUL-1997 (TrEMBLrel. 04, Created)  
DE 01-JUL-1997 (TrEMBLrel. 04, Last sequence update)  
DE 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
DE Poly(ADP-ribose) glycohydrolase.  
GN BPARG.  
OS Bos taurus (Bovine).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
OC Bovidae; Bovinae; Bos.  
OX NCBI\_TaxID=9913;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC MEDLINE=92727328; PubMed=9115250;  
RA Lin W.; Ame J.C.; Aboul-Ela N.; Jacobson E.L.; Jacobson M.K.;  
RT "Isolation and characterization of the cDNA encoding bovine poly(ADP-  
RT ribose) glycohydrolase";  
RL J. Biol. Chem. 272:11895-11901 (1997).  
DR EMBL; U78975; AAH53370.1; -;  
DR GO; GO:0016787; F:hydrolase activity; IEA.  
DR InterPro; IPR007724; PARG.  
DR Pfam; PF05028; PARG; 1.  
KW Hydrolase.  
SQ SEQUENCE 977 AA; 110837 MW; 87D2100F979DF377 CRC64;  
  
Query Match 100.0%; Score 167; DB 6; Length 977;  
Best Local Similarity 100.0%; Pred. No. 2.7e-15;  
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 LFTVLDHNECLITGTQYSEYTGATYR 31  
DB 771 LFTVLDHNECLITGTQYSEYTGATYR 801  
|||||  
[1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Berkley;  
RA Stapleton M.; Brokstein P.; Hong L.; Agbayani A.; Carlson J.;  
RA Champe M.; Chavez C.; Dorsett V.; Farfan D.; Frise E.; George R.;  
RA Gonzalez M.; Guarin H.; Li P.; Liao G.; Miranda A.; Mungall C.J.;  
RA Nunoo J.; Pacleb J.; Paragas V.; Park S.; Phouanavong S.; Wan K.;  
RA Yu C.; Lewis S.E.; Rubin G.M.; Celniker S.;  
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.

```

DR EMBL; AY051955; AAK93379.1; -.
DR FlyBase; FBgn0023216; PARG.
DR InterPro; IPR007724; PARG.
DR Pfam; PF05028; PARG; 1.
SQ SEQUENCE 723 AA; 81109 MW; 90A5BA59775D611C CRC64;

Query Match
Best Local Similarity 49.1%; Score 82; DB 5; Length 723;
Matches 16; Conservative 5; Mismatches 9; Indels 0; Gaps 0;

QY 1 LFTEVLHDNECLITGTQYSEYTGAEYTY 30
DB 355 LFTECLRPFEALVMLGAERYSNYTGAGSF 384

RESULT 10
O46043 PRELIMINARY; PRT; 768 AA.
AC O46043;
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE PARG protein.
GN PARG OR EG:114E2.1 OR CG2864.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkeley;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celisner S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Vandeil M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers J.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abriel J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berland P.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotier P.,
RA Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Fostler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kianos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Scapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195(2000).
RN [2]
RP SEQUENCE FROM N.A.

```

```

RA Ame J.-C., Jacobson M.K.;
RT "Isolation and characterization of the cDNA encoding Drosophila
RT poly(ADP-ribose) glycohydrolase."
RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 46-768 FROM N.A.
RA Murphy L., Harris D., Barrell B.;
RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE OF 46-768 FROM N.A.
RA Benos P.;
RL Submitted (DEC-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AE003428; AAF45896.1; -.
DR EMBL; AF079556; AAC28734.1; -.
DR EMBL; Z98254; CAB10913.1; -.
DR FlyBase; FBgn0023216; PARG.
DR GO; GO:0016787; F:hydrolase activity; IEA.
DR InterPro; IPR007724; PARG.
DR Pfam; PF05028; PARG; 1.
KW Hydrolase.
SQ SEQUENCE 768 AA; 86288 MW; 2E620C42E643FC04 CRC64;

Query Match
Best Local Similarity 49.1%; Score 82; DB 5; Length 768;
Matches 16; Conservative 5; Mismatches 9; Indels 0; Gaps 0;

QY 1 LFTEVLHDNECLITGTQYSEYTGAEYTY 30
DB 400 LFTECLRPFEALVMLGAERYSNYTGAGSF 429

RESULT 11
Q9SKB3 PRELIMINARY; PRT; 548 AA.
AC Q9SKB3; Q9AET7;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Putative poly(ADP-ribose) glycohydrolase.
GN ATG31870 OR TEJ.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsi.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RA Lin X., Kaul S., Shea T.P., Fujii C.Y., Shen M., VanAken S.E.,
RA Barnstead M.E., Mason T.M., Bowman C.L., Ronning C.M., Benito M.-I.,
RA Carrera A.J., Creasy T.H., Buell C.R., Town C.D., Nierman W.C.,
RA Fraser C.M., Venter J.C.;
RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Panda S., Poirier G.G., Kay S.A.;
RT "TEJ defines a role for poly-ADP-riboseylation in establishing period
RT length of the Arabidopsis circadian oscillator."
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC006533; AAD32285.2; -.
DR EMBL; AF394690; AAK72256.1; -.
DR PIR; B84726; B84726.
DR GO; GO:0016787; F:hydrolase activity; IEA.
DR InterPro; IPR007724; PARG.
DR Pfam; PF05028; PARG; 1.
KW Hydrolase.
SQ SEQUENCE 548 AA; 62169 MW; F1A79FDA157C3329 CRC64;

Query Match
Best Local Similarity 47.3%; Score 79; DB 10; Length 548;
Matches 15; Conservative 6; Mismatches 10; Indels 0; Gaps 0;

QY 1 LFTEVLHDNECLITGTQYSEYTGAEYTY 31

```

Matches 9; Conservative 11; Mismatches 11; Indels 0; Gaps 0

Db 288 LFLPRDDNEAIEIVGAERFSCYTGYSFPR 318

RESULT 12

Q8VYAI PRELIMINARY; PRT; 522 AA.

ID Q8VYAI

AC Q8VYAI

DT 01-MAR-2002 (TREMELrel. 20, Created)

DT 01-MAR-2002 (TREMELrel. 20, Last sequence update)

DT 01-JUN-2003 (TREMELrel. 24, Last annotation update)

DT Putative poly(ADP-ribose) glycohydrolase.

GN AT2G31870

OS Arabidopsis thaliana (Mouse-ear cress).

OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;

OC eurosids II; Brassicales; Brassicaceae; Arabidopsi.

OC NCBI\_TaxID=3702;

OX [1]

RN SEQUENCE FROM N.A.

RA Nguyen M., Karlin-Neumann G., Southwick A., Lam B., Miranda M.,

RA Palm C.J., Bowser L., Jones T., Banh J., Carninci P., Chen H.,

RA Cheuk R., Chung M.K., Hayashizaki Y., Ishida J., Kamiya A., Kawai J.,

RA Kim C., Lin J., Liu S.X., Narusaka M., Pham P.K., Sakano H.,

RA Sakurai T., Satou M., Seki M., Shinn P., Yamada K., Shinozaki K.,

RA Ecker J., Theologis A., Davis R.W.;

RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.

DR EMBL; AY072330; AAL61937.1; --

DR GO; GO:0016787; F:hydrolase activity; IEA.

DR InterPro; IPR007724; PARG.

DR Pfam; PF05028; PARG; 1.

KW Hydrolase.

SQ SEQUENCE 522 AA; 58942 MW; 6PF98685D867CA15 CRC64;

Query Match 41.3%; Score 69; DB 10; Length 522;

Best Local Similarity 38.7%; Pred. No. 0.2;

Matches 12; Conservative 8; Mismatches 11; Indels 0; Gaps 0;

QY 1 LFTVELDHNELIITGTEQYSEYTGVAETPR 31

Db 299 IFLPRDDNEAIEIVGAERFSCYTGYSFQ 329

RESULT 13

Q86G14 PRELIMINARY; PRT; 368 AA.

ID Q86G14

AC Q86G14

DT 01-JUN-2003 (TREMELrel. 24, Created)

DT 01-JUN-2003 (TREMELrel. 24, Last sequence update)

DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)

DE Non-transmembrane antigen.

OS Toxoplasma gondii

OC Eukaryota; Alveolata; Apicomplexa; Coccidia; Eimeriida; Sarcocystidae;

OC Toxoplasma.

OC NCBI\_TaxID=5811;

OX [1]

RN SEQUENCE FROM N.A.

RA Cai L., Wang D., Shu H., Jiang L., Wu X., Zeng Q.;

RA "Immunological screening of Toxoplasma tachyzoite cDNA expression

RT libraries with serum from toxoplasma infected rabbits.;"

RL Submitted (JAN-2003) to the EMBL/GenBank/DBJ databases.

DR EMBL; AY223538; AAC65977.1; --

DR GO; GO:0016021; C:integral to membrane; IEA.

DR GO; GO:0005215; F:transporter activity; IEA.

DR GO; GO:0006810; P:transport; IEA.

DR InterPro; IPR005829; Sug transporter.

DR PROSITE; PS00216; SUGAR\_TRANSPORT\_1; 1.

KW Transmembrane.

SQ SEQUENCE 368 AA; 40914 MW; 7751A41095E6549C CRC64;

Query Match 34.1%; Score 57; DB 5; Length 368;

Best Local Similarity 29.0%; Pred. No. 7.5;

Matches 11; Conservative 5; Mismatches 15; Indels 0; Gaps 0;

QY 1 LFTVELDHNELIITGTEQYSEYTGVAETPR 31

Db 286 LINDVTQDLAISIVGAERFSCYTGYSNLT 316

RESULT 15

Q8TR71 PRELIMINARY; PRT; 437 AA.

ID Q8TR71

AC Q8TR71

DT 01-JUN-2002 (TREMELrel. 21, Created)

DT 01-JUN-2002 (TREMELrel. 21, Last sequence update)

DT 01-JUN-2002 (TREMELrel. 21, Last annotation update)

DE Hypothetical protein MA1312.

GN MA1312.

OS Methanosarcina acetivorans.

Query Match 32.9%; Score 55; DB 5; Length 485;

Best Local Similarity 35.5%; Pred. No. 20;

Matches 11; Conservative 5; Mismatches 15; Indels 0; Gaps 0;

QY 1 LFTVELDHNELIITGTEQYSEYTGVAETPR 31

Db 286 LINDVTQDLAISIVGAERFSCYTGYSNLT 316

OC Archaea: Euryarchaeota; Euryarchaeota orders incertae sedis;  
 OC Methanosarcinales; Methanosarcinaceae; Methanosarcina.  
 OX NCBI\_taxid=2214;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C2A / ATCC 35395 / DSM 2834;  
 RX MEDLINE=21929760; PubMed=11932238;  
 RA Galagan J.E., Nusbbaum C., Roy A., Endrizzi M.G., Macdonald P.,  
 RA FitzHugh W., Calvo S., Engels-Thumann N., DeArelano K., Johnson R.,  
 RA Allen N., Naylor J., Stange-Thomann N., DeArelano K., Johnson R.,  
 RA Linton L., McEwan P., McKernan K., Talamas J., Tirrell A., Ye W.,  
 RA Zimmer A., Barber R.D., Cann I., Graham D.E., Grahame D.A., Guss A.M.,  
 RA Hedderich R., Ingram-Smith C., Kuetner H.C., Krzycki J.A.,  
 RA Leigh J.A., Li W., Liu J., Mukhopadhyay B., Reeve J.N., Smith K.,  
 RA Springer T.A., Umayam L.A., White O., White R.H., de Macario E.C.,  
 RA Ferry J.G., Jarrell K.F., Jing H., Macario A.J.B., Paulsen I.,  
 RA Pritchett M., Sowers K.E., Swanson R.V., Zinder S.H., Lander E.,  
 RA Metcalf W.W., Birren B.;  
 RT "The genome of Methanosarcina acetivorans reveals extensive metabolic  
 RT and physiological diversity.";   
 RL Genome Res. 12:532-542(2002).  
 DR EMBL; AE010800; AAM04729.1; --  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 437 AA; 48229 MW; BB4A089F86593F5E CRC64;

Query Match 32.0%; Score 53.5; DB 17; Length 437;  
 Best Local Similarity 46.2%; Pred. No. 29;  
 Matches 12; Conservative 5; Mismatches 8; Indels 1; Gaps 1;

QY 1 LFTVEVDHNECLITGTQYSEYTG 26  
 :|||||:|||||:  
 Db 172 VFEEVLD-ADLIISTGTVEFHYAGY 196

Search completed: May 26, 2004, 18:46:08  
 Job time : 4.474 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 26, 2004, 17:19:25 ; Search time 3.32255 Seconds  
(without alignments)  
2466.140 Million cell updates/sec

Title: US-09-302-812-12

Perfect score: 152  
Sequence: 1 AYCGLRPGVSSSENLSAVATGNKGCAGF 29

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_29Jan04:\*

- 1: Geneseqp1980s:\*
- 2: Geneseqp1990s:\*
- 3: Geneseqp2000s:\*
- 4: Geneseqp2001s:\*
- 5: Geneseqp2002s:\*
- 6: Geneseqp2003as:\*
- 7: Geneseqp2003bs:\*
- 8: Geneseqp2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	150	98.7	29	5 AAE25635	Bovine po
2	150	98.7	29	5 AAU76026	Bovine po
3	150	98.7	29	5 AAU76017	Bovine po
4	150	98.7	29	6 ABG72284	Oligopept
5	149	98.0	976	5 AAE25630	Human pol
6	149	98.0	976	5 AAU76021	Human pol
7	149	98.0	976	5 AAU76012	Human pol
8	149	98.0	976	6 ABG72279	Human pol
9	149	98.0	977	5 AAE25629	Bovine po
10	149	98.0	977	5 AAU76020	Bovine po
11	149	98.0	977	5 AAU75799	Bovine po
12	149	98.0	977	6 ABG72278	Bovine po
13	144	94.7	33	5 AAE25652	Bovine po
14	144	94.7	968	5 AAE25631	Murine po
15	144	94.7	968	5 AAU76022	Mouse pol
16	144	94.7	968	5 AAU76013	Mouse pol
17	144	94.7	968	6 ABG72280	Murine p
18	69	45.4	768	4 ABB59491	Drosophil
19	69	45.4	768	5 AAE25632	Fruit fly
20	69	45.4	768	5 AAU76023	Fruit fly
21	69	45.4	768	5 AAU76014	Fruit fly
22	69	45.4	768	6 ABG72281	Fruit fly
23	60.5	39.8	543	2 AAR42456	Enzyme in
24	60.5	39.8	543	2 AAR39465	Biosynthe
25	60.5	39.8	543	2 AAW37053	S. putref

26	60.5	39.8	543	2 AAW89403	S. putref
27	60.5	39.8	543	3 AAB10470	Shewanell
28	55	36.2	578	4 ABB71910	Drosophil
29	53	34.9	75	5 ABU67233	G-protein
30	53	34.9	75	6 ABP54014	Human neu
31	53	34.9	403	3 AAY44642	Human gro
32	53	34.9	403	3 AAY90673	Human mut
33	53	34.9	403	3 AAY90638	Human G p
34	53	34.9	403	4 AAB99185	Human FM-
35	53	34.9	403	5 ABG95158	Human GPC
36	53	34.9	403	5 ABG95171	Human GPC
37	53	34.9	403	6 ABP81682	Human G p
38	53	34.9	403	6 ADA99010	Human NMU
39	53	34.9	403	7 ADC22833	Human G p
40	53	34.9	403	7 ADC22743	Human G p
41	53	34.9	403	7 ADE31653	Human 143
42	53	34.9	426	4 AAR03628	Human G-p
43	53	34.9	426	5 AAE15631	Human G-p
44	53	34.9	445	4 AAU68566	Human nov
45	53	34.9	445	4 AAU68523	Human nov

ALIGNMENTS

RESULT 1  
AAE25635  
ID AAE25635 standard; peptide; 29 AA.

XX AAE25635;

XX 04-NOV-2002 (first entry)

XX Bovine poly adenosine diphosphate-ribose glycohydrolase oligopeptide #2.  
XX Bovine: poly adenosine diphosphate-ribose glycohydrolase; bPARG; enzyme;  
KW ADP-ribose; neoplastic disorder; genotoxic oxidative stress; nootropic;  
KW cardiac disorder; neuronal disorder; reperfusion injury; neurotoxicity;  
KW Alzheimer's disease; Huntington's disease; Parkinson's disease; cardiac;  
KW cytostatic; vasotropic; neuroprotective; anticonvulsant; gene therapy;  
KW antisense therapy.

XX Bos taurus.

XX Key Location/Qualifiers  
FH Misc-difference 23  
FT /label= Unknown

XX US6395543-B1.

XX 28-MAY-2002.

XX 23-FEB-2000; 2000US-00511507.

XX 01-MAY-1998; 98US-0083768P.

XX 30-APR-1999; 99US-00302812.

XX (KENT ) UNIV KENTUCKY RES FOUND.

XX Jacobson MK, Jacobson BL, Ame J, Lin W;

XX WPI; 2002-535641/57.

XX New nucleic acid molecule encoding bovine poly adenosine diphosphate-ribose glycohydrolase involved in cellular response to DNA damage, inhibition of which is useful for treating neoplastic disorders and neurodegenerative diseases.

XX Claim 10; Col 81-82; 77pp; English.

XX The invention relates to an isolated nucleic acid molecule which encodes a bovine poly (adenosine diphosphate (ADP)-ribose) glycohydrolase (bPARG) which catalyses release of ADP-ribose from an ADP-ribose polymer. The

CC invention is useful as probes and primer molecules that can be used in  
 CC hybridisation assays and polymerase chain reaction (PCR) amplification.  
 CC The knowledge of the nucleotide sequence of the PARG gene permits the  
 CC preparation of antisense therapeutics containing sequences complementary  
 CC to the mRNA of PARG gene. The antisense therapeutic are useful to treat  
 CC neoplastic disorders and conditions caused by genotoxic oxidative stress  
 CC e.g., cardiac disorders, neuronal disorders, reperfusion injury,  
 CC neurotoxicity, Alzheimer's disease, Huntington's disease and Parkinson's  
 CC disease. The invention is useful in gene therapy and antisense therapy.  
 CC The present sequence is bPARG oligopeptide  
 XX  
 XX Sequence 29 AA;  
 SQ

Query Match 98.7%; Score 150; DB 5; Length 29;  
 Best Local Similarity 100.0%; Pred. No. 3.2e-14; Length 29;  
 Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AYCGLRPGVSSSENLSAVATGNKXGCGAFG 29  
 |||||  
 DB 1 AYCGLRPGVSSSENLSAVATGNKXGCGAFG 29  
 |||||

RESULT 3  
 AAU76026  
 ID AAU76026 standard; peptide; 29 AA.  
 XX  
 AC AAU76026;  
 XX  
 DT 08-MAY-2002 (first entry)  
 XX  
 DE Bovine poly(ADP-ribose) glycohydrolase (PARG) oligopeptide 63.  
 XX  
 KW Bovine; poly(ADP-ribose) glycohydrolase; PARG; DNA repair; ADP-ribose;  
 KW adenosine diphosphate-ribose; neoplastic disorder; adenocarcinoma;  
 KW leukaemia; lymphoma; melanoma; myeloma; sarcoma; teratocarcinoma;  
 KW hyperplasia; hypertrophy; reperfusion; ischaemia; heart attack; stroke;  
 KW neurodegenerative disease; neurological disorder; Alzheimer's disease;  
 KW Huntington's disease; Parkinson's disease; oligopeptide 63.  
 XX  
 OS Bos taurus.  
 XX

Key Location/Qualifiers  
 FH Misc-difference 23  
 FT /label= Unknown  
 FT /note= 'Xaa is not further defined in the specification'  
 FT  
 PN US6337202-B1.  
 XX  
 PD 08-JAN-2002.  
 XX  
 PF 23-FEB-2000; 2000US-00511477.  
 XX  
 PR 01-MAY-1998; 98US-0083768P.  
 XX  
 PR 30-APR-1999; 99US-00302812.  
 XX  
 PA (KENT ) UNIV KENTUCKY RES FOUND.  
 XX  
 PI Jacobson MK, Jacobson EL, Ame J, Lin W;  
 XX  
 DR WPI; 2002-163240/21.  
 XX  
 PT Novel isolated and purified poly (ADP-ribose) glycohydrolase protein  
 PT which catalyses release of ADP-ribose from ADP ribose polymer, useful for  
 PT treating neoplastic and neurological disorders, heart attack and stroke.  
 XX  
 PS Claim 4; Col 25; 81pp; English.  
 XX  
 CC The present invention relates to a new poly(ADP-ribose) glycohydrolase  
 CC (PARG) protein which catalyses release of ADP-ribose from an ADP  
 CC (adenosine diphosphate)-ribose polymer. The PARG molecule of the  
 CC invention is useful for generating antibodies and can be inhibited or  
 CC activated for diagnosing and treating neoplastic disorders such as  
 CC adenocarcinoma, leukaemia, lymphoma, melanoma, myeloma, sarcoma,

CC teratocarcinoma, hyperplasia and hypertrophy, reperfusion following  
 CC ischaemia, heart attack, stroke, neurodegenerative diseases, neurological  
 CC disorders including Alzheimer's, Huntington's and Parkinson's diseases,  
 CC and related conditions. PARG levels may be enhanced to suppress DNA  
 CC repair and increase the cell's susceptibility to chemotherapy drugs.  
 CC Antagonists of PARG are administered to treat or prevent neoplastic  
 CC disorders. The present amino acid sequence represents bovine PARG  
 CC oligonucleotide 63. This peptide is one of several PARG oligopeptides  
 CC (AAU76025-AAU76028) of the invention  
 XX  
 XX Sequence 29 AA;  
 SQ

Query Match 98.7%; Score 150; DB 5; Length 29;  
 Best Local Similarity 100.0%; Pred. No. 3.2e-14; Length 29;  
 Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AYCGLRPGVSSSENLSAVATGNKXGCGAFG 29  
 |||||  
 DB 1 AYCGLRPGVSSSENLSAVATGNKXGCGAFG 29  
 |||||

RESULT 3  
 AAU76017  
 ID AAU76017 standard; peptide; 29 AA.  
 XX  
 AC AAU76017;  
 XX  
 DT 08-MAY-2002 (first entry)  
 XX  
 DE Bovine poly(ADP-ribose) glycohydrolase (PARG) oligopeptide 63.  
 XX  
 KW Cow; poly(ADP-ribose) glycohydrolase; PARG; PARG modulator; ADP-ribose;  
 KW adenosine diphosphate-ribose; DNA repair; apoptosis; neoplasia; cancer;  
 KW inherited genetic disease; myocardial infarction; vascular stroke; aging;  
 KW neurodegeneration; Huntington's disease; Parkinson's disease;  
 KW Alzheimer's disease; neurotoxicity; oligopeptide 63.  
 XX  
 OS Bos taurus.  
 XX

Key Location/Qualifiers  
 FH Misc-difference 23  
 FT /label= Unknown  
 FT /note= 'Xaa is not further defined in the specification'  
 FT  
 PN US6333148-B1.  
 XX  
 PD 25-DEC-2001.  
 XX  
 PF 30-APR-1999; 99US-00302812.  
 XX  
 PF 01-MAY-1998; 98US-0083768P.  
 XX  
 PR (KENT ) UNIV KENTUCKY RES FOUND.  
 XX  
 PI Jacobson MK, Jacobson EL, Ame J, Lin W;  
 XX  
 DR WPI; 2002-153820/20.  
 XX  
 PT Screening compounds for modulation of poly(ADP-ribose) glycohydrolase,  
 PT useful potentially for treating diseases associated with DNA damage, e.g.  
 PT cancer.  
 XX  
 PS Claim 4; Col 25; 80pp; English.  
 XX  
 CC The present invention relates to a new method for screening compounds for  
 CC ability to modulate activity of an enzyme that hydrolyses ADP (adenosine  
 CC diphosphate)-ribose from an ADP-ribose polymer. The compounds are  
 CC inhibitors or activators of PARG (poly(ADP-ribose) glycohydrolase) and  
 CC are used to treat or prevent any condition associated with DNA damage  
 CC e.g. neoplasia, inherited genetic diseases, myocardial infarction,  
 CC Parkinson's disease, aging and neurodegeneration e.g. Huntington's,  
 CC Parkinson's or Alzheimer's diseases, or neurotoxicity generally.  
 CC Compounds identified by the new method are more effective than known

CC inhibitors and have fewer side effects. The present amino acid sequence  
 CC represents bovine PARG oligopeptide 63. This peptide is one of several  
 CC PARG oligopeptides (AAU76016-AAU76019) of the invention  
 XX  
 XX Sequence 29 AA;  
 SQ

Query Match 98.7%; Score 150; DB 5; Length 29;  
 Best Local Similarity 100.0%; Pred. No. 3.2e-14;  
 Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AYCGLRPGVSSNLSAVATGNXGCGAFG 29  
 DB 1 AYCGLRPGVSSNLSAVATGNXGCGAFG 29  
 |||||

RESULT 4  
 ABG72284  
 ID ABG72284 standard; peptide; 29 AA.  
 XX  
 AC ABG72284;  
 XX  
 DT 13-MAR-2003 (first entry)  
 XX  
 DE Oligopeptide #2 derived from bovine PARG enzyme.  
 XX  
 KW Bovine; poly(ADP-ribose) glycohydrolase; PARG; PARG expression;  
 KW cellular response; DNA damage; neoplastic disorder inducing agent;  
 KW oxidative stress; neoplastic disorder; myocardial infarction;  
 KW vascular stroke; neurodegenerative disease; Alzheimer's disease;  
 KW Parkinson's disease; Huntington's disease; inborn genetic error;  
 KW reperfusion; ischemia; aging; neurotoxicity; PARG activity; cytostatic;  
 KW neuroprotective; nontropic; antiparkinsonian; cardiant; vasotropic;  
 KW anticonvulsant; cerebroprotective.  
 XX  
 OS Bos taurus.  
 XX

Key Location/Qualifiers  
 FH Misc-difference 23  
 FT /label= Unknown  
 FT  
 XX US2002132328-A1.  
 XX  
 XX 19-SEP-2002.  
 XX  
 XX 09-OCT-2001; 2001US-00973451.  
 XX  
 XX 01-MAY-1998; 98US-0083768P.  
 XX  
 XX 30-APR-1999; 99US-00302812.  
 XX  
 XX (JACO/) JACOBSON M K.  
 XX (JACO/) JACOBSON E L.  
 XX (AMEJ/) AME J.  
 XX (LINW/) LIN W.  
 XX  
 XX Jacobson MK, Jacobson EL, Ame J, Lin W;  
 PI  
 XX WPI; 2003-155895/15.  
 XX  
 XX New nucleic acid molecule encoding a polypeptide with poly(ADP-ribose)  
 PT glycohydrolase activity, for preventing, treating, or ameliorating a  
 PT disease condition, e.g. neoplastic disorder, myocardial infarction or  
 PT vascular stroke.  
 FT  
 XX  
 XX Example 2; Page 14; 86pp; English.  
 PS  
 XX The present invention relates to the isolation of poly(ADP-ribose)  
 XX glycohydrolase (PARG) from several species, and the polynucleotide  
 CC sequences encoding them. Methods for inhibiting PARG expression or  
 CC overexpressing PARG are also disclosed. PARG is involved in the cellular  
 CC response to DNA damage, and is associated with the body's response to  
 CC neoplastic disorder inducing agents and oxidative stress. The  
 CC polynucleotide sequences encoding PARG and PARG modulators are useful for  
 CC preventing, treating, or ameliorating diseases such as neoplastic

CC disorders, myocardial infarction, vascular stroke, neurodegenerative  
 CC disorders (e.g. Alzheimer's disease, Parkinson's disease or Huntington's  
 CC disease), inborn genetic errors, reperfusion following ischemia, aging,  
 CC and neurotoxicity. The polynucleotide sequences are also useful in gene  
 CC therapy. The methods are useful for identifying an agent that modulates  
 CC PARG activity, identifying a mutant PARG allele in an individual, or  
 CC screening candidate molecules for PARG modulating activity. ABG72283-  
 CC ABG72286 represent oligopeptides derived from bovine PARG enzyme. The  
 CC oligopeptides are used to construct degenerate PCR primers for the  
 CC isolation of cDNA encoding bovine PARG  
 XX  
 XX Sequence 29 AA;  
 SQ

Query Match 98.7%; Score 150; DB 6; Length 29;  
 Best Local Similarity 100.0%; Pred. No. 3.2e-14;  
 Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AYCGLRPGVSSNLSAVATGNXGCGAFG 29  
 DB 1 AYCGLRPGVSSNLSAVATGNXGCGAFG 29  
 |||||

RESULT 5  
 AAE25630  
 ID AAE25630 standard; protein; 976 AA.  
 XX  
 AC AAE25630;  
 XX  
 DT 04-NOV-2002 (first entry)  
 XX  
 DE Human poly adenosine diphosphate-ribose glycohydrolase (PARG).  
 XX  
 KW Human; poly adenosine diphosphate-ribose glycohydrolase; PARG; enzyme;  
 KW ADP-ribose; neoplastic disorder; genotoxic oxidative stress; nontropic;  
 KW cardiac disorder; neuronal disorder; reperfusion injury; neurotoxicity;  
 KW Alzheimer's disease; Huntington's disease; Parkinson's disease; cardiant;  
 KW cytostatic; vasotropic; neuroprotective; anticonvulsant; gene therapy;  
 KW antisense therapy.  
 XX  
 XX Homo sapiens.  
 XX  
 XX US6395543-B1.  
 XX  
 XX 28-MAY-2002.  
 XX  
 XX 23-FEB-2000; 2000US-00511507.  
 XX  
 XX 01-MAY-1998; 98US-0083768P.  
 XX  
 XX 30-APR-1999; 99US-00302812.  
 XX  
 XX (KENT ) UNIV KENTUCKY RES FOUND.  
 XX  
 XX Jacobson MK, Jacobson EL, Ame J, Lin W;  
 PI  
 XX WPI; 2002-535641/57.  
 XX  
 XX N-PSDB; AAD42082.  
 XX  
 XX New nucleic acid molecule encoding bovine poly adenosine diphosphate-  
 PT ribose glycohydrolase involved in cellular response to DNA damage,  
 PT inhibition of which is useful for treating neoplastic disorders and  
 PT neurodegenerative diseases.  
 XX  
 XX Claim 3; Col 55-60; 77pp; English.  
 PS  
 XX The invention relates to an isolated nucleic acid molecule which encodes  
 CC a bovine poly (adenosine diphosphate (ADP)-ribose) glycohydrolase (BPARG)  
 CC which catalyses release of ADP-ribose from an ADP ribose polymer. The  
 CC invention is useful as probes and primer molecules that can be used in  
 CC hybridization assays and polymerase chain reaction (PCR) amplification.  
 CC The knowledge of the nucleotide sequence of the PARG gene permits the  
 CC preparation of antisense therapeutics containing sequences complementary  
 CC to the mRNA of PARG gene. The antisense therapeutic are useful to treat  
 CC neoplastic disorders and conditions caused by genotoxic oxidative stress

CC e.g., cardiac disorders, neuronal disorders, reperfusion injury,  
 CC neurotoxicity, Alzheimer's disease, Huntington's disease and Parkinson's  
 CC disease. The invention is useful in gene therapy and antisense therapy.  
 CC The present sequence is human PARG  
 XX  
 SQ Sequence 976 AA;

Query Match 98.0%; Score 149; DB 5; Length 976;  
 Best Local Similarity 96.6%; Pred. No. 1.8e-12;  
 Matches 28; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AYCGLRPGVSSSENLSAVATGNXGCGAFG 29  
 |||||  
 Db 848 AYCGLRPGVSSSENLSAVATGNXGCGAFG 876  
 |||||

RESULT 6  
 AAU76021  
 ID AAU76021 standard; protein; 976 AA.  
 XX  
 AC AAU76021;

XX  
 DT 08-MAY-2002 (first entry)  
 XX  
 DE Human poly(ADP-ribose) glycohydrolase (PARG) protein sequence.

XX Human; poly(ADP-ribose) glycohydrolase; PARG; DNA repair; ADP-ribose;  
 KW adenosine diphosphate-ribose; neoplastic disorder; adenocarcinoma;  
 KW leukaemia; lymphoma; melanoma; myeloma; sarcoma; teratocarcinoma;  
 KW hyperplasia; hypertrophy; reperfusion; ischaemia; heart attack; stroke;  
 KW neurodegenerative disease; neurological disorder; Alzheimer's disease;  
 KW Huntington's disease; Parkinson's disease.

XX Homo sapiens.  
 XX US6337202-B1.  
 XX 08-JAN-2002.

XX 23-FEB-2000; 2000US-00511477.  
 XX 01-MAY-1998; 98US-0083768P.  
 XX 30-APR-1999; 99US-00302812.

XX (KENT ) UNIV KENTUCKY RES FOUND.  
 XX Jacobson MK, Jacobson EL, Ame J, Lin W;  
 XX WPI; 2002-163240/21.  
 XX N-PSDB; ABK14932.

XX Novel isolated and purified poly (ADP-ribose) glycohydrolase protein  
 PT which catalyses release of ADP-ribose from ADP ribose polymer, useful for  
 PT treating neoplastic and neurological disorders, heart attack and stroke.  
 XX  
 PS Claim 2; Col 55-60; 81pp; English.

XX The present invention relates to a new poly(ADP-ribose) glycohydrolase  
 CC (PARG) protein which catalyses release of ADP-ribose from an ADP  
 CC (adenosine diphosphate)-ribose polymer. The PARG molecule of the  
 CC invention is useful for generating antibodies and can be inhibited or  
 CC activated for diagnosing and treating neoplastic disorders such as  
 CC adenocarcinoma, leukaemia, lymphoma, melanoma, myeloma, sarcoma,  
 CC teratocarcinoma, hyperplasia and hypertrophy, reperfusion following  
 CC ischaemia, heart attack, stroke, neurodegenerative diseases, neurological  
 CC disorders including Alzheimer's, Huntington's and Parkinson's diseases,  
 CC and related conditions. PARG levels may be enhanced to suppress DNA  
 CC repair and increase the cell's susceptibility to chemotherapy drugs.  
 CC Antagonists of PARG are administered to treat or prevent neoplastic  
 CC disorders. The present amino acid sequence represents the human PARG  
 CC protein of the invention. This protein is one of several PARG proteins  
 CC (AAU76020-AAU76024) of the invention

SQ Sequence 976 AA;  
 Query Match 98.0%; Score 149; DB 5; Length 976;  
 Best Local Similarity 96.6%; Pred. No. 1.8e-12;  
 Matches 28; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AYCGLRPGVSSSENLSAVATGNXGCGAFG 29  
 |||||  
 Db 848 AYCGLRPGVSSSENLSAVATGNXGCGAFG 876  
 |||||

RESULT 7  
 AAU76012  
 ID AAU76012 standard; protein; 976 AA.  
 XX  
 AC AAU76012;

XX  
 DT 08-MAY-2002 (first entry)  
 XX  
 DE Human poly(ADP-ribose) glycohydrolase (PARG) protein sequence.

XX Human; poly(ADP-ribose) glycohydrolase; PARG; PARG modulator; ADP-ribose;  
 KW adenosine diphosphate-ribose; DNA repair; apoptosis; neoplasia; cancer;  
 KW inherited genetic disease; myocardial infarction; vascular stroke; aging;  
 KW neurodegeneration; Huntington's disease; Parkinson's disease;  
 KW Alzheimer's disease; neurotoxicity.

XX Homo sapiens.  
 XX US6333148-B1.  
 XX 25-DEC-2001.

XX 30-APR-1999; 99US-00302812.  
 XX 01-MAY-1998; 98US-0083768P.  
 XX (KENT ) UNIV KENTUCKY RES FOUND.

XX Jacobson MK, Jacobson EL, Ame J, Lin W;  
 XX WPI; 2002-153820/20.  
 XX N-PSDB; ABK14494.

XX Screening compounds for modulation of poly(ADP-ribose) glycohydrolase,  
 PT useful potentially for treating diseases associated with DNA damage, e.g.  
 PT cancer.  
 XX  
 PS Claim 3; Col 55-60; 80pp; English.

XX The present invention relates to a new method for screening compounds for  
 CC ability to modulate activity of an enzyme that hydrolyses ADP (adenosine  
 CC diphosphate)-ribose from an ADP-ribose polymer. The compounds are  
 CC inhibitors or activators of PARG (poly(ADP-ribose) glycohydrolase) and  
 CC are used to treat or prevent any condition associated with DNA damage,  
 CC e.g. neoplasia, inherited genetic diseases, myocardial infarction,  
 CC vascular stroke, aging and neurodegeneration e.g. Huntington's,  
 CC Parkinson's or Alzheimer's diseases, or neurotoxicity generally.  
 CC Compounds identified by the new method are more effective than known  
 CC inhibitors and have fewer side effects. The present amino acid sequence  
 CC represents the human PARG protein of the invention. This protein is one  
 CC of several PARG proteins (AAU75799 and AAU76012-AAU76015) of the  
 CC invention

SQ Sequence 976 AA;  
 Query Match 98.0%; Score 149; DB 5; Length 976;  
 Best Local Similarity 96.6%; Pred. No. 1.8e-12;  
 Matches 28; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AYCGLRPGVSSSENLSAVATGNXGCGAFG 29  
 |||||  
 Db 848 AYCGLRPGVSSSENLSAVATGNXGCGAFG 876  
 |||||

RESULT 8  
 ABG72279  
 ID ABG72279 standard; protein; 976 AA.  
 XX AC ABG72279;  
 XX DT 13-MAR-2003 (first entry)  
 XX DE Human poly(ADP-ribose) glycohydrolase (PARG) enzyme.  
 XX KW Human; poly(ADP-ribose) glycohydrolase; PARG; PARG expression;  
 KW cellular response; DNA damage; neoplastic disorder inducing agent;  
 KW oxidative stress; neoplastic disorder; myocardial infarction;  
 KW vascular stroke; neurodegenerative disease; Alzheimer's disease;  
 KW Parkinson's disease; Huntington's disease; inborn genetic error;  
 KW reperfusion; ischemia; aging; neurotoxicity; PARG activity; cytostatic;  
 KW neuroprotective; neotropic; antiparkinsonian; cardiant; vasotropic;  
 KW anticonvulsant; cerebroprotective; enzyme.  
 XX OS Homo sapiens.  
 XX PN US2002132328-A1.  
 XX PD 19-SEP-2002.  
 XX PF 09-OCT-2001; 2001US-00973451.  
 XX PR 01-MAY-1998; 98US-0083768P.  
 XX PR 30-APR-1999; 99US-00302812.  
 XX PA (JACO/) JACOBSON M K.  
 XX PA (JACO/) JACOBSON E L.  
 XX PA (AMEJ/) AME J.  
 XX PA (LINW/) LIN W.  
 XX PI Jacobson MK, Jacobson EL, Ame J, Lin W;  
 XX WPI; 2003-155895/15.  
 XX DR N-PSDB; ABX14478.  
 XX PT New nucleic acid molecule encoding a polypeptide with poly(ADP-ribose)  
 PT glycohydrolase activity, for preventing, treating, or ameliorating a  
 PT disease condition, e.g. neoplastic disorder, myocardial infarction or  
 PT vascular stroke.  
 XX PS Claim 28; Fig 16; 86pp; English.  
 XX CC The present invention relates to the isolation of poly(ADP-ribose)  
 CC glycohydrolase (PARG) from several species, and the polynucleotide  
 CC sequences encoding them. Methods for inhibiting PARG expression or  
 CC overexpressing PARG are also disclosed. PARG is involved in the cellular  
 CC response to DNA damage, and is associated with the body's response to  
 CC neoplastic disorder inducing agents and oxidative stress. The  
 CC polynucleotide sequences encoding PARG and PARG modulators are useful for  
 CC preventing, treating, or ameliorating diseases such as neoplastic  
 CC disorders, myocardial infarction, vascular stroke, neurodegenerative  
 CC disorders (e.g. Alzheimer's disease, Parkinson's disease or Huntington's  
 CC disease), inborn genetic errors, reperfusion following ischaemia, aging,  
 CC and neurotoxicity. The polynucleotide sequences are also useful in gene  
 CC therapy. The methods are useful for identifying an agent that modulates  
 CC PARG activity, identifying a mutant PARG allele in an individual, or  
 CC screening candidate molecules for PARG modulating activity. The present  
 CC sequence represents human PARG enzyme  
 XX SQ Sequence 976 AA;  
 Query Match 98.0%; Score 149; DB 6; Length 976;  
 Best Local Similarity 96.6%; Pred. No. 1.8e-12;  
 Matches 28; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 AYCGFLRPGVSSNLSAVATGNXGCGAGF 29

Db 848 AYCGFLRPGVSSNLSAVATGNXGCGAGF 876  
 |||||  
 RESULT 9  
 AAE25629  
 ID AAE25629 standard; protein; 977 AA.  
 XX AC AAE25629;  
 XX DT 04-NOV-2002 (first entry)  
 XX DE Bovine poly adenosine diphosphate-ribose glycohydrolase (bPARG).  
 XX KW Bovine; poly adenosine diphosphate-ribose glycohydrolase; bPARG; enzyme;  
 KW ADP-ribose; neoplastic disorder; genotoxic oxidative stress; neotropic;  
 KW cardiac disorder; neuronal disorder; reperfusion injury; neurotoxicity;  
 KW Alzheimer's disease; Huntington's disease; Parkinson's disease; cardiant;  
 KW cytostatic; vasotropic; neuroprotective; anticonvulsant; gene therapy;  
 KW antisense therapy.  
 XX OS Bos taurus.  
 XX PN US6395543-B1.  
 XX PD 28-MAY-2002.  
 XX PF 23-FEB-2000; 2000US-00511507.  
 XX PR 01-MAY-1998; 98US-0083768P.  
 XX PR 30-APR-1999; 99US-00302812.  
 XX PA (KENT ) UNIV KENTUCKY RES FOUND.  
 XX PI Jacobson MK, Jacobson EL, Ame J, Lin W;  
 XX WPI; 2002-535641/57.  
 XX DR N-PSDB; AAD42081.  
 XX PT New nucleic acid molecule encoding bovine poly adenosine diphosphate-  
 PT ribose glycohydrolase involved in cellular response to DNA damage,  
 PT inhibition of which is useful for treating neoplastic disorders and  
 PT neurodegenerative diseases.  
 XX PS Claim 3; Col 47-45; 77pp; English.  
 XX CC The invention relates to an isolated nucleic acid molecule which encodes  
 CC a bovine poly (adenosine diphosphate (ADP)-ribose) glycohydrolase (bPARG)  
 CC which catalyses release of ADP-ribose from an ADP-ribose polymer. The  
 CC invention is useful as probes and primer molecules that can be used in  
 CC hybridisation assays and polymerase chain reaction (PCR) amplification.  
 CC The knowledge of the nucleotide sequence of the PARG gene permits the  
 CC preparation of antisense therapeutics containing sequences complementary  
 CC to the mRNA of PARG gene. The antisense therapeutic are useful to treat  
 CC neoplastic disorders and conditions caused by genotoxic oxidative stress  
 CC e.g., cardiac disorders, neuronal disorders, reperfusion injury,  
 CC neurotoxicity, Alzheimer's disease, Huntington's disease and Parkinson's  
 CC disease. The invention is useful in gene therapy and antisense therapy.  
 XX SQ Sequence 977 AA;  
 Query Match 98.0%; Score 149; DB 5; Length 977;  
 Best Local Similarity 96.6%; Pred. No. 1.8e-12;  
 Matches 28; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 AYCGFLRPGVSSNLSAVATGNXGCGAGF 29  
 |||||  
 Db 849 AYCGFLRPGVSSNLSAVATGNXGCGAGF 877  
 |||||  
 RESULT 10  
 AAU76020

ID XX AAU76020 standard; protein; 977 AA.  
AC XX AAU76020;  
XX XX  
DT XX 08-MAY-2002 (first entry)  
XX XX  
DE XX Bovine poly (ADP-ribose) glycohydrolase (PARG) protein sequence.  
XX XX  
KW Cow; poly (ADP-ribose) glycohydrolase; PARG; DNA repair; ADP-ribose;  
KW adenosine diphosphate-ribose; neoplastic disorder; adenocarcinoma;  
KW leukaemia; lymphoma; melanoma; myeloma; sarcoma; teratocarcinoma;  
KW hyperplasia; hypertrophy; reperfusion; ischaemia; heart attack; stroke;  
KW neurodegenerative disease; neurological disorder; Alzheimer's disease;  
KW Huntington's disease; Parkinson's disease.  
XX XX  
OS Bos taurus.  
XX XX  
FH Key Location/Qualifiers  
FT Region 601..617  
FT /note= "Represents PARG oligopeptide #1"  
FT Region 761..770  
FT /note= "Represents PARG oligopeptide #2"  
FT Region 771..801  
FT /note= "Represents PARG oligopeptide #3"  
FT Region 849..880  
FT /note= "Represents PARG oligopeptide #4"  
XX XX  
PN USG337202-B1.  
XX XX  
PD 08-JAN-2002.  
XX XX  
PF 23-FEB-2000; 2000US-00511477.  
XX XX  
PR 01-MAY-1998; 98US-0083768P.  
PR 30-APR-1999; 99US-00302812.  
XX XX  
PA (KENT ) UNIV KENTUCKY RES FOUND.  
XX XX  
PI Jacobson MK, Jacobson EL, Ame J, Lin W;  
XX WPI; 2002-163240/21.  
DR N-PSDB; ABK14931.  
XX XX  
PT Novel isolated and purified poly (ADP-ribose) glycohydrolase protein  
PT which catalyses release of ADP-ribose from ADP ribose polymer, useful for  
PT treating neoplastic and neurological disorders, heart attack and stroke.  
XX XX  
PS Claim 2; Col 47-52; 81pp; English.  
XX XX  
CC The present invention relates to a new poly (ADP-ribose) glycohydrolase  
CC (PARG) protein which catalyses release of ADP-ribose from an ADP  
CC (adenosine diphosphate)-ribose polymer. The PARG molecule of the  
CC invention is useful for generating antibodies and can be inhibited or  
CC activated for diagnosing and treating neoplastic disorders such as  
CC adenocarcinoma, leukaemia, lymphoma, melanoma, myeloma, sarcoma,  
CC teratocarcinoma, hyperplasia and hypertrophy, reperfusion following  
CC ischaemia, heart attack, stroke, neurodegenerative diseases, neurological  
CC disorders including Alzheimer's, Huntington's and Parkinson's diseases,  
CC and related conditions. PARG levels may be enhanced to suppress DNA  
CC repair and increase the cell's susceptibility to chemotherapy drugs.  
CC Antagonists of PARG are administered to treat or prevent neoplastic  
CC disorders. The present amino acid sequence represents the bovine PARG  
CC protein of the invention. This protein is one of several PARG proteins  
CC (AAU76020-AAU76024) of the invention  
XX XX  
SQ Sequence 977 AA;

Query Match 98.0%; Score 149; DB 5; Length 977;  
Best Local Similarity 96.6%; Pred. No. 1.8e-12;  
Matches 28; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 1 AYCGLRPGVSSSENLSAVATGNXGCGAFG 29  
|||||

Db 849 AYCGLRPGVSSSENLSAVATGNXGCGAFG 877  
RESULT 11  
AAU75799  
ID AAU75799 standard; protein; 977 AA.  
XX AC AAU75799;  
XX XX  
DT XX 08-MAY-2002 (first entry)  
XX XX  
DE XX Bovine poly (ADP-ribose) glycohydrolase (PARG) protein sequence.  
XX XX  
KW Cow; poly (ADP-ribose) glycohydrolase; PARG; PARG modulator; ADP-ribose;  
KW adenosine diphosphate-ribose; DNA repair; apoptosis; neoplasia; cancer;  
KW inherited genetic disease; myocardial infarction; vascular stroke; aging;  
KW neurodegeneration; Huntington's disease; Parkinson's disease;  
KW Alzheimer's disease; neurotoxicity.  
XX XX  
OS Bos taurus.  
XX XX  
FH Key Location/Qualifiers  
FT Region 601..617  
FT /note= "Represents PARG oligopeptide #1"  
FT Region 761..770  
FT /note= "Represents PARG oligopeptide #2"  
FT Region 771..801  
FT /note= "Represents PARG oligopeptide #3"  
FT Region 849..880  
FT /note= "Represents PARG oligopeptide #4"  
XX XX  
PN US6333148-B1.  
XX XX  
PD 25-DEC-2001.  
XX XX  
PF 30-APR-1999; 99US-00302812.  
XX XX  
PR 01-MAY-1998; 98US-0083768P.  
XX XX  
PI (KENT ) UNIV KENTUCKY RES FOUND.  
XX XX  
PI Jacobson MK, Jacobson EL, Ame J, Lin W;  
XX WPI; 2002-153820/20.  
DR N-PSDB; ABK14493.  
XX XX  
PT Screening compounds for modulation of poly (ADP-ribose) glycohydrolase,  
PT useful potentially for treating diseases associated with DNA damage, e.g.  
PT cancer.  
XX XX  
PS Claim 3; Col 45-52; 80pp; English.  
XX XX  
CC The present invention relates to a new method for screening compounds for  
CC ability to modulate activity of an enzyme that hydrolyses ADP (adenosine  
CC diphosphate)-ribose from an ADP-ribose polymer. The compounds are  
CC inhibitors or activators of PARG (poly (ADP-ribose) glycohydrolase) and  
CC are used to treat or prevent any condition associated with DNA damage  
CC e.g. neoplasia, inherited genetic diseases, myocardial infarction,  
CC vascular stroke, aging and neurodegeneration e.g. Huntington's,  
CC Parkinson's or Alzheimer's diseases, or neurotoxicity generally.  
CC Compounds identified by the new method are more effective than known  
CC inhibitors and have fewer side effects. The present amino acid sequence  
CC represents the bovine PARG protein of the invention. This protein is one  
CC of several PARG proteins (AAU75799 and AAU76012-AAU76015) of the  
CC invention  
XX XX  
SQ Sequence 977 AA;

Query Match 98.0%; Score 149; DB 5; Length 977;  
Best Local Similarity 96.6%; Pred. No. 1.8e-12;  
Matches 28; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 1 AYCGLRPGVSSSENLSAVATGNXGCGAFG 29

Db 849 AYCGLRPGVSSSENLSAVATGNWCGGAFG 877  
|||||

## RESULT 12

ABG72278  
ID ABG72278 standard; protein; 977 AA.

XX AC ABG72278;  
XX DT 13-MAR-2003 (first entry)

XX DE Bovine poly (ADP-ribose) glycohydrolase (PARG) enzyme.  
XX KW Bovine; poly (ADP-ribose) glycohydrolase; PARG; PARG expression;  
KW cellular response; DNA damage; neoplastic disorder inducing agent;  
KW oxidative stress; neoplastic disorder; myocardial infarction;  
KW vascular stroke; neurodegenerative disorder; Alzheimer's disease;  
KW Parkinson's disease; Huntington's disease; inborn genetic error;  
KW reperfusion; ischaemia; aging; neurotoxicity; PARG activity; cytostatic;  
KW neuroprotective; neotropic; antiparkinsonian; cardiant; vasotropic;  
KW anticonvulsant; cerebroprotective; enzyme.

XX OS Bos taurus.  
XX PN US2002132328-A1.  
XX PD 19-SEP-2002.

XX PF 09-OCT-2001; 2001US-00973451.  
XX PR 01-MAY-1998; 98US-0083768P.  
XX PR 30-APR-1999; 99US-00302812.

XX PA (JACO/) JACOBSON M K.  
XX PA (JACO/) JACOBSON E L.  
XX PA (AMEJ/) AME J.  
XX PA (LINW/) LIN W.

XX PI Jacobson MK, Jacobson EL, Ame J, Lin W;  
XX WPI; 2003-155895/15.  
XX DR N-PSDB; ABX14477.

XX PT New nucleic acid molecule encoding a polypeptide with poly (ADP-ribose) glycohydrolase activity, for preventing, treating, or ameliorating a disease condition, e.g. neoplastic disorder, myocardial infarction or vascular stroke.  
XX PS Claim 28; Fig 16; 86pp; English.  
XX CC The present invention relates to the isolation of poly (ADP-ribose) glycohydrolase (PARG) from several species, and the polynucleotide sequences encoding them. Methods for inhibiting PARG expression or overexpressing PARG are also disclosed. PARG is involved in the cellular response to DNA damage, and is associated with the body's response to neoplastic disorder inducing agents and oxidative stress. The polynucleotide sequences encoding PARG and PARG modulators are useful for preventing, treating, or ameliorating diseases such as neoplastic disorders, myocardial infarction, vascular stroke, neurodegenerative disorders (e.g. Alzheimer's disease, Parkinson's disease or Huntington's disease), inborn genetic errors, reperfusion following ischaemia, aging, and neurotoxicity. The polynucleotide sequences are also useful in gene therapy. The methods are useful for identifying an agent that modulates PARG activity, identifying a mutant PARG allele in an individual, or screening candidate molecules for PARG modulating activity. The present sequence represents bovine PARG enzyme

XX SQ Sequence 977 AA;  
XX Query Match 98.0%; Score 149; DB 6; Length 977;  
XX Best Local Similarity 96.6%; Pred. No. 1.8e-12;  
XX Matches 28; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AYCGLRPGVSSSENLSAVATGNKGGCGAFG 29  
|||||

Db 849 AYCGLRPGVSSSENLSAVATGNWCGGAFG 877  
|||||

## RESULT 13

AAE25652  
ID AAE25652 standard; peptide; 33 AA.

XX AC AAE25652;  
XX DT 04-NOV-2002 (first entry)

XX DE Bovine poly adenosine diphosphate-ribose glycohydrolase oligopeptide #8.  
XX KW Bovine; poly adenosine diphosphate-ribose glycohydrolase; bPARG; enzyme;  
KW ADP-ribose; neoplastic disorder; genotoxic oxidative stress; neurotoxic;  
KW cardiac disorder; neuronal disorder; reperfusion injury; neurotoxicity;  
KW Alzheimer's disease; Huntington's disease; Parkinson's disease; cardiant;  
KW cytostatic; vasotropic; neuroprotective; anticonvulsant; gene therapy;  
KW antisense therapy.

XX OS Bos taurus.  
XX PN US6395543-B1.  
XX PD 28-MAY-2002.

XX PF 23-FEB-2000; 2000US-00511507.  
XX PR 01-MAY-1998; 98US-0083768P.  
XX PR 30-APR-1999; 99US-00302812.

XX PA (KENT ) UNIV KENTUCKY RES FOUND.  
XX PI Jacobson MK, Jacobson EL, Ame J, Lin W;  
XX DR WPI; 2002-535641/57.

XX PT New nucleic acid molecule encoding bovine poly adenosine diphosphate-ribose glycohydrolase involved in cellular response to DNA damage, inhibition of which is useful for treating neoplastic disorders and neurodegenerative diseases.  
XX PS Example 3; Col 27; 77pp; English.  
XX CC The invention relates to an isolated nucleic acid molecule which encodes a bovine poly (adenosine diphosphate (ADP)-ribose) glycohydrolase (bPARG) which catalyses release of ADP-ribose from an ADP-ribose polymer. The invention is useful as probes and primer molecules that can be used in hybridisation assays and polymerase chain reaction (PCR) amplification. The knowledge of the nucleotide sequence of the PARG gene permits the preparation of antisense therapeutics containing sequences complementary to the mRNA of PARG gene. The antisense therapeutic are useful to treat neoplastic disorders and conditions caused by genotoxic oxidative stress e.g., cardiac disorders, neuronal disorders, reperfusion injury, neurotoxicity, Alzheimer's disease, Huntington's disease and Parkinson's disease. The invention is useful in gene therapy and antisense therapy. The present sequence is bPARG oligopeptide

XX SQ Sequence 33 AA;  
XX Query Match 94.7%; Score 144; DB 5; Length 33;  
XX Best Local Similarity 93.1%; Pred. No. 2.6e-13;  
XX Matches 27; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AYCGLRPGVSSSENLSAVATGNKGGCGAFG 29  
|||||

Db 1 AYCGLRPGVSSSENLSAVATGNWCGGAFG 29  
|||||

## RESULT 14

AAE25631  
 ID AAE25631 standard; protein; 968 AA.  
 AC AAE25631;  
 XX  
 DT 04-NOV-2002 (first entry)  
 DE Murine poly adenosine diphosphate-ribose glycohydrolase (PARG).  
 XX  
 XX Murine: poly adenosine diphosphate-ribose glycohydrolase; PARG; enzyme;  
 KW ADP-ribose; neoplastic disorder; genotoxic oxidative stress; neutropenic;  
 KW cardiac disorder; neuronal disorder; reperfusion injury; neurotoxicity;  
 KW Alzheimer's disease; Huntington's disease; Parkinson's disease; cardiac;  
 KW cytostatic; vasotropic; neuroprotective; anticonvulsant; gene therapy;  
 KW antisense therapy.  
 XX  
 OS Mus musculus.  
 XX  
 FN US6395543-B1.  
 XX  
 PD 28-MAY-2002.  
 XX  
 PF 23-FEB-2000; 2000US-00511507.  
 XX  
 PR 01-MAY-1998; 98US-0083768P.  
 PR 30-APR-1999; 99US-00302812.  
 XX  
 PA (KENT ) UNIV KENTUCKY RES FOUND.  
 XX  
 PI Jacobson MK, Jacobson EL, Ame J, Lin W;  
 XX  
 DR WPI; 2002-535641/57.  
 DR N-PSDB; AAD42083.  
 XX  
 XX New nucleic acid molecule encoding bovine poly adenosine diphosphate-  
 PT ribose glycohydrolase involved in cellular response to DNA damage,  
 PT inhibition of which is useful for treating neoplastic disorders and  
 PT neurodegenerative diseases.  
 XX  
 PS Claim 3; Col 63-68; 77pp; English.  
 CC  
 CC The invention relates to an isolated nucleic acid molecule which encodes  
 CC a bovine poly (adenosine diphosphate (ADP)-ribose) glycohydrolase (bPARG)  
 CC which catalyses release of ADP-ribose from an ADP ribose polymer. The  
 CC invention is useful as probes and primer molecules that can be used in  
 CC hybridisation assays and polymerase chain reaction (PCR) amplification.  
 CC The knowledge of the nucleotide sequence of the PARG gene permits the  
 CC preparation of antisense therapeutics containing sequences complementary  
 CC to the mRNA of PARG gene. The antisense therapeutic are useful to treat  
 CC neoplastic disorders and conditions caused by genotoxic oxidative stress  
 CC e.g., cardiac disorders, neuronal disorders, reperfusion injury,  
 CC neurotoxicity, Alzheimer's disease, Huntington's disease and Parkinson's  
 CC disease. The invention is useful in gene therapy and antisense therapy.  
 CC The present sequence is murine PARG  
 XX  
 SQ Sequence 968 AA;  
 Query Match 94.7%; Score 144; DB 5; Length 968;  
 Best Local Similarity 93.1%; Pred. No. 9.2e-12;  
 Matches 27; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 Qy 1 AYCGLRPGVSENLSAVATGNKCGAGF 29  
 |||||  
 Db 840 AYCGLRPGVSENLSAVATGNKCGAGF 868  
 |||||  
 RESULT 15  
 AAU76022  
 ID AAU76022 standard; protein; 968 AA.  
 XX  
 AC AAU76022;  
 XX  
 DT 08-MAY-2002 (first entry)

XX  
 DE Mouse poly (ADP-ribose) glycohydrolase (PARG) protein sequence.  
 XX  
 KW Mouse: poly (ADP-ribose) glycohydrolase; PARG; DNA repair; ADP-ribose;  
 KW adenosine diphosphate-ribose; neoplastic disorder; adenocarcinoma;  
 KW leukaemia; lymphoma; melanoma; myeloma; sarcoma; teratocarcinoma;  
 KW hyperplasia; hypertrophy; reperfusion; ischaemia; heart attack; stroke;  
 KW neurodegenerative disease; neurological disorder; Alzheimer's disease;  
 KW Huntington's disease; Parkinson's disease.  
 OS  
 XX Mus musculus.  
 XX  
 FN US6337202-B1.  
 XX  
 PD 08-JAN-2002.  
 XX  
 PF 23-FEB-2000; 2000US-00511477.  
 XX  
 PR 01-MAY-1998; 98US-0083768P.  
 PR 30-APR-1999; 99US-00302812.  
 XX  
 PA (KENT ) UNIV KENTUCKY RES FOUND.  
 XX  
 PI Jacobson MK, Jacobson EL, Ame J, Lin W;  
 XX  
 DR WPI; 2002-163240/21.  
 DR N-PSDB; ABK14933.  
 XX  
 PT Novel isolated and purified poly (ADP-ribose) glycohydrolase protein  
 PT which catalyses release of ADP-ribose from ADP ribose polymer, useful for  
 PT treating neoplastic and neurological disorders, heart attack and stroke.  
 XX  
 PS Claim 2; Col 63-70; 81pp; English.  
 CC  
 CC The present invention relates to a new poly (ADP-ribose) glycohydrolase  
 CC (PARG) protein which catalyses release of ADP-ribose from an ADP  
 CC (adenosine diphosphate)-ribose polymer. The PARG molecule of the  
 CC invention is useful for generating antibodies and can be inhibited or  
 CC activated for diagnosing and treating neoplastic disorders such as  
 CC adenocarcinoma, leukaemia, lymphoma, melanoma, myeloma, sarcoma,  
 CC teratocarcinoma, hyperplasia and hypertrophy, reperfusion following  
 CC ischaemia, heart attack, stroke, neurodegenerative diseases, neurological  
 CC disorders including Alzheimer's, Huntington's and Parkinson's diseases,  
 CC repair and increase the cell's susceptibility to chemotherapy drugs.  
 CC Antagonists of PARG are administered to treat or prevent neoplastic  
 CC disorders. The present amino acid sequence represents the mouse PARG  
 CC protein of the invention. This protein is one of several PARG proteins  
 CC (AAU76020-AAU76024) of the invention  
 XX  
 SQ Sequence 968 AA;  
 Query Match 94.7%; Score 144; DB 5; Length 968;  
 Best Local Similarity 93.1%; Pred. No. 9.2e-12;  
 Matches 27; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 Qy 1 AYCGLRPGVSENLSAVATGNKCGAGF 29  
 |||||  
 Db 840 AYCGLRPGVSENLSAVATGNKCGAGF 868  
 |||||

Search completed: May 26, 2004, 18:40:09  
 Job time : 4.32255 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 26, 2004, 18:38:05 ; Search time 0.929262 Seconds  
(without alignments)  
1611.120 Million cell updates/sec

Title: US-09-302-812-12  
Perfect score: 152  
Sequence: 1 AYCGLRPGVSSSENLSAVATGNKGCAGF 29

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents AA:\*  
1: /cgn2\_6/prodata/2/iaa/5A COMB.pep.\*  
2: /cgn2\_6/prodata/2/iaa/5B COMB.pep.\*  
3: /cgn2\_6/prodata/2/iaa/6A COMB.pep.\*  
4: /cgn2\_6/prodata/2/iaa/6B COMB.pep.\*  
5: /cgn2\_6/prodata/2/iaa/PCTUS COMB.pep.\*  
6: /cgn2\_6/prodata/2/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	150	98.7	29	4	US-09-302-812-12
2	150	98.7	29	4	US-09-302-812-12
3	150	98.7	29	4	US-09-302-812-12
4	149	98.0	976	4	US-09-302-812-4
5	149	98.0	976	4	US-09-302-812-4
6	149	98.0	976	4	US-09-302-812-4
7	149	98.0	977	4	US-09-302-812-2
8	149	98.0	977	4	US-09-302-812-2
9	149	98.0	977	4	US-09-302-812-2
10	144	94.7	968	4	US-09-302-812-6
11	144	94.7	968	4	US-09-302-812-6
12	144	94.7	968	4	US-09-302-812-6
13	69	45.4	768	4	US-09-302-812-8
14	69	45.4	768	4	US-09-302-812-8
15	69	45.4	768	4	US-09-302-812-8
16	60.5	39.8	543	1	US-08-375-709-17
17	60.5	39.8	543	1	US-08-375-709-17
18	60.5	39.8	543	3	US-09-090-793-10
19	60.5	39.8	543	4	US-09-231-899-10
20	53	34.9	403	4	US-09-170-496D-114
21	53	34.9	403	4	US-09-170-496D-114
22	53	34.9	403	4	US-09-743-742B-4
23	53	34.9	403	4	US-09-743-742B-10
24	50.5	33.2	411	4	US-09-540-236-3549
25	49.5	32.6	726	4	US-09-302-812-10
26	49.5	32.6	726	4	US-09-302-812-10
27	49.5	32.6	726	4	US-09-302-812-10

28	49	32.2	1196	4	US-09-252-991A-18261	Sequence 18261, A
29	48.5	31.9	887	4	US-09-077-940A-2	Sequence 2, Appli
30	47.5	31.2	298	3	US-09-025-691-1	Sequence 1, Appli
31	46.5	30.6	134	4	US-09-621-976-4066	Sequence 4066, Ap
32	46.5	30.6	307	3	US-09-049-672A-9	Sequence 9, Appli
33	46	30.3	359	2	US-08-713-636-2	Sequence 2, Appli
34	46	30.3	681	1	US-08-083-590A-18	Sequence 18, Appl
35	46	30.3	681	2	US-08-346-128-37	Sequence 37, Appl
36	46	30.3	681	3	US-08-532-384-18	Sequence 18, Appl
37	46	30.3	1068	1	US-08-537-210A-2	Sequence 2, Appli
38	46	30.3	1068	3	US-09-113-825-2	Sequence 2, Appli
39	46	30.3	1078	1	US-08-264-534-32	Sequence 32, Appl
40	46	30.3	1078	1	US-08-083-590A-11	Sequence 11, Appl
41	46	30.3	1078	1	US-08-465-500-32	Sequence 32, Appl
42	46	30.3	1078	2	US-08-346-128-32	Sequence 32, Appl
43	46	30.3	1078	3	US-08-532-384-11	Sequence 11, Appl
44	46	30.3	1078	3	US-08-893-828-32	Sequence 32, Appl
45	46	30.3	2556	1	US-08-185-432-17	Sequence 17, Appl

ALIGNMENTS

RESULT 1  
US-09-302-812-12  
; Sequence 12, Application US/09302812B  
; Patent No. 6333148.  
; GENERAL INFORMATION:  
; APPLICANT: JACOBSON, Myron K.  
; APPLICANT: JACOBSON, Elaine L.  
; APPLICANT: AM, Jean-Christophe  
; APPLICANT: LIN, Winston  
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE (PARG) ENZ  
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIVE  
; TITLE OF INVENTION: THEREWITH  
; FILE REFERENCE: NIAD 201  
; CURRENT APPLICATION NUMBER: US/09/302,812B  
; CURRENT FILING DATE: 1999-04-30  
; EARLIER APPLICATION NUMBER: 60/083,768  
; EARLIER FILING DATE: 1998-05-01  
; NUMBER OF SEQ ID NOS: 38  
; SEQ ID NO 12  
; LENGTH: 29  
; TYPE: PRT  
; ORGANISM: Bos taurus  
; FEATURE:  
US-09-302-812-12

Query Match 98.7%; Score 150; DB 4; Length 29;  
Best Local Similarity 100.0%; Pred. No. 1.2e-14;  
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AYCGLRPGVSSSENLSAVATGNKGCAGF 29  
DB 1 AYCGLRPGVSSSENLSAVATGNKGCAGF 29

RESULT 2  
US-09-511-477-12  
; Sequence 12, Application US/09511477  
; Patent No. 6337202  
; GENERAL INFORMATION:  
; APPLICANT: JACOBSON, Myron K.  
; APPLICANT: AME, Jean-Christophe  
; APPLICANT: LIN, Winston  
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE (PARG) ENZ  
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIVE  
; TITLE OF INVENTION: THEREWITH  
; FILE REFERENCE: NIAD 201  
; CURRENT APPLICATION NUMBER: US/09/511,477  
; CURRENT FILING DATE: 2000-02-23  
; PRIOR APPLICATION NUMBER: 09/302,812

; PRIOR FILING DATE: 1999-04-30  
; NUMBER OF SEQ ID NOS: 38  
; SEQ ID NO 12  
; LENGTH: 29  
; TYPE: PRT  
; ORGANISM: Bos taurus  
; FEATURE:  
US-09-511-477-12

Query Match 98.7%; Score 150; DB 4; Length 29;  
Best Local Similarity 100.0%; Pred. No. 1.2e-14;  
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AYCGLRPGVSSSENLSAVATGNKGCAGF 29  
|||  
DB 1 AYCGLRPGVSSSENLSAVATGNKGCAGF 29  
|||

## RESULT 3

US-09-511-507-12  
; Sequence 12, Application US/09511507  
; Patent No. 6395543  
; GENERAL INFORMATION:  
; APPLICANT: JACOBSON, Myron K.  
; APPLICANT: JACOBSON, Elaine L.  
; APPLICANT: AME, Jean-Christophe  
; APPLICANT: LIN, Winston  
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY(ADP-RIBOSE) GLYCOHYDROLASE (PARG) EN  
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV  
; TITLE OF INVENTION: THEREWITH  
; FILE REFERENCE: NIAD 201  
; CURRENT APPLICATION NUMBER: US/09/511,507  
; CURRENT FILING DATE: 2000-02-23  
; PRIOR APPLICATION NUMBER: 09/302,812  
; PRIOR FILING DATE: 1999-04-30  
; NUMBER OF SEQ ID NOS: 38  
; SEQ ID NO 12  
; LENGTH: 29  
; TYPE: PRT  
; ORGANISM: Bos taurus  
; FEATURE:  
US-09-511-507-12

Query Match 98.7%; Score 150; DB 4; Length 29;  
Best Local Similarity 100.0%; Pred. No. 1.2e-14;  
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AYCGLRPGVSSSENLSAVATGNKGCAGF 29  
|||  
DB 1 AYCGLRPGVSSSENLSAVATGNKGCAGF 29  
|||

## RESULT 4

US-09-302-812-4  
; Sequence 4, Application US/09302812B  
; Patent No. 6333148  
; GENERAL INFORMATION:  
; APPLICANT: JACOBSON, Myron K.  
; APPLICANT: JACOBSON, Elaine L.  
; APPLICANT: AM, Jean-Christophe  
; APPLICANT: LIN, Winston  
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY(ADP-RIBOSE) GLYCOHYDROLASE (PARG) EN  
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV  
; TITLE OF INVENTION: THEREWITH  
; FILE REFERENCE: NIAD 201  
; CURRENT APPLICATION NUMBER: US/09/302,812B  
; CURRENT FILING DATE: 1999-04-30  
; EARLIER APPLICATION NUMBER: 60/083,768  
; EARLIER FILING DATE: 1998-05-01  
; NUMBER OF SEQ ID NOS: 38  
; SEQ ID NO 4  
; LENGTH: 976  
; TYPE: PRT

; ORGANISM: Homo sapiens  
; FEATURE:  
US-09-302-812-4

Query Match 98.0%; Score 149; DB 4; Length 976;  
Best Local Similarity 96.6%; Pred. No. 7e-13;  
Matches 28; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AYCGLRPGVSSSENLSAVATGNKGCAGF 29  
|||  
DB 848 AYCGLRPGVSSSENLSAVATGNKGCAGF 876  
|||

## RESULT 5

US-09-511-477-4  
; Sequence 4, Application US/09511477  
; Patent No. 6337202  
; GENERAL INFORMATION:  
; APPLICANT: JACOBSON, Myron K.  
; APPLICANT: JACOBSON, Elaine L.  
; APPLICANT: AME, Jean-Christophe  
; APPLICANT: LIN, Winston  
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY(ADP-RIBOSE) GLYCOHYDROLASE (PARG) EN  
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV  
; TITLE OF INVENTION: THEREWITH  
; FILE REFERENCE: NIAD 201  
; CURRENT APPLICATION NUMBER: US/09/511,477  
; CURRENT FILING DATE: 2000-02-23  
; PRIOR APPLICATION NUMBER: 09/302,812  
; PRIOR FILING DATE: 1999-04-30  
; NUMBER OF SEQ ID NOS: 38  
; SEQ ID NO 4  
; LENGTH: 976  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
US-09-511-477-4

Query Match 98.0%; Score 149; DB 4; Length 976;  
Best Local Similarity 96.6%; Pred. No. 7e-13;  
Matches 28; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AYCGLRPGVSSSENLSAVATGNKGCAGF 29  
|||  
DB 848 AYCGLRPGVSSSENLSAVATGNKGCAGF 876  
|||

## RESULT 6

US-09-511-507-4  
; Sequence 4, Application US/09511507  
; Patent No. 6395543  
; GENERAL INFORMATION:  
; APPLICANT: JACOBSON, Myron K.  
; APPLICANT: JACOBSON, Elaine L.  
; APPLICANT: AME, Jean-Christophe  
; APPLICANT: LIN, Winston  
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY(ADP-RIBOSE) GLYCOHYDROLASE (PARG) EN  
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV  
; TITLE OF INVENTION: THEREWITH  
; FILE REFERENCE: NIAD 201  
; CURRENT APPLICATION NUMBER: US/09/511,507  
; CURRENT FILING DATE: 2000-02-23  
; PRIOR APPLICATION NUMBER: 09/302,812  
; PRIOR FILING DATE: 1999-04-30  
; NUMBER OF SEQ ID NOS: 38  
; SEQ ID NO 4  
; LENGTH: 976  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
US-09-511-507-4

Query Match 98.0%; Score 149; DB 4; Length 976;

Best Local Similarity 96.6%; Pred. No. 7e-13; Indels 1; Mismatches 0; Gaps 0;  
Matches 28; Conservative 0;

QY 1 AYCGLRPGVSSNLSAVATGNXGCGAFG 29  
|||||  
Db 848 AYCGLRPGVSSNLSAVATGNXGCGAFG 876  
|||||

RESULT 7  
US-09-302-812-2  
; Sequence 2, Application US/09302812B  
; Patent No. 6333148  
; GENERAL INFORMATION:  
; APPLICANT: JACOBSON, Myron K.  
; APPLICANT: JACOBSON, Elaine L.  
; APPLICANT: AME, Jean-Christophe  
; APPLICANT: LIN, Winston  
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE (PARG) ENZ  
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIVE  
; TITLE OF INVENTION: THEREWITH  
; FILE REFERENCE: NIAD 201  
; CURRENT APPLICATION NUMBER: US/09/302,812B  
; CURRENT FILING DATE: 1999-04-30  
; EARLIER APPLICATION NUMBER: 60/083,768  
; EARLIER FILING DATE: 1998-05-01  
; NUMBER OF SEQ ID NOS: 38  
; SEQ ID NO 2  
; LENGTH: 977  
; TYPE: PRT  
; ORGANISM: Bos taurus  
; FEATURE:  
US-09-302-812-2

Query Match 98.0%; Score 149; DB 4; Length 977;  
Best Local Similarity 96.6%; Pred. No. 7e-13; Indels 1; Mismatches 0; Gaps 0;  
Matches 28; Conservative 0;

QY 1 AYCGLRPGVSSNLSAVATGNXGCGAFG 29  
|||||  
Db 849 AYCGLRPGVSSNLSAVATGNXGCGAFG 877  
|||||

RESULT 8  
US-09-511-477-2  
; Sequence 2, Application US/09511477  
; Patent No. 6337202  
; GENERAL INFORMATION:  
; APPLICANT: JACOBSON, Myron K.  
; APPLICANT: JACOBSON, Elaine L.  
; APPLICANT: AME, Jean-Christophe  
; APPLICANT: LIN, Winston  
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE (PARG) ENZ  
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIVE  
; TITLE OF INVENTION: THEREWITH  
; FILE REFERENCE: NIAD 201  
; CURRENT APPLICATION NUMBER: US/09/511,477  
; CURRENT FILING DATE: 2000-02-23  
; PRIOR APPLICATION NUMBER: 09/302,812  
; PRIOR FILING DATE: 1999-04-30  
; NUMBER OF SEQ ID NOS: 38  
; SEQ ID NO 2  
; LENGTH: 977  
; TYPE: PRT  
; ORGANISM: Bos taurus  
; FEATURE:  
US-09-511-477-2

Query Match 98.0%; Score 149; DB 4; Length 977;  
Best Local Similarity 96.6%; Pred. No. 7e-13; Indels 1; Mismatches 0; Gaps 0;  
Matches 28; Conservative 0;

QY 1 AYCGLRPGVSSNLSAVATGNXGCGAFG 29  
|||||

Db 849 AYCGLRPGVSSNLSAVATGNXGCGAFG 877  
|||||

RESULT 9  
US-09-511-507-2  
; Sequence 2, Application US/09511507  
; Patent No. 6395543  
; GENERAL INFORMATION:  
; APPLICANT: JACOBSON, Myron K.  
; APPLICANT: JACOBSON, Elaine L.  
; APPLICANT: AME, Jean-Christophe  
; APPLICANT: LIN, Winston  
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE (PARG) ENZ  
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIVE  
; TITLE OF INVENTION: THEREWITH  
; FILE REFERENCE: NIAD 201  
; CURRENT APPLICATION NUMBER: US/09/511,507  
; CURRENT FILING DATE: 2000-02-23  
; PRIOR APPLICATION NUMBER: 09/302,812  
; PRIOR FILING DATE: 1999-04-30  
; NUMBER OF SEQ ID NOS: 38  
; SEQ ID NO 2  
; LENGTH: 977  
; TYPE: PRT  
; ORGANISM: Bos taurus  
; FEATURE:  
US-09-511-507-2

Query Match 98.0%; Score 149; DB 4; Length 977;  
Best Local Similarity 96.6%; Pred. No. 7e-13; Indels 1; Mismatches 0; Gaps 0;  
Matches 28; Conservative 0;

QY 1 AYCGLRPGVSSNLSAVATGNXGCGAFG 29  
|||||  
Db 849 AYCGLRPGVSSNLSAVATGNXGCGAFG 877  
|||||

RESULT 10  
US-09-302-812-6  
; Sequence 6, Application US/09302812B  
; Patent No. 6333148  
; GENERAL INFORMATION:  
; APPLICANT: JACOBSON, Myron K.  
; APPLICANT: JACOBSON, Elaine L.  
; APPLICANT: AM, Jean-Christophe  
; APPLICANT: LIN, Winston  
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE (PARG) ENZ  
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIVE  
; TITLE OF INVENTION: THEREWITH  
; FILE REFERENCE: NIAD 201  
; CURRENT APPLICATION NUMBER: US/09/302,812B  
; CURRENT FILING DATE: 1999-04-30  
; EARLIER APPLICATION NUMBER: 60/083,768  
; EARLIER FILING DATE: 1998-05-01  
; NUMBER OF SEQ ID NOS: 38  
; SEQ ID NO 6  
; LENGTH: 968  
; TYPE: PRT  
; ORGANISM: Mus musculus  
; FEATURE:  
US-09-302-812-6

Query Match 94.7%; Score 144; DB 4; Length 968;  
Best Local Similarity 93.1%; Pred. No. 3.5e-12; Indels 2; Mismatches 0; Gaps 0;  
Matches 27; Conservative 0;

QY 1 AYCGLRPGVSSNLSAVATGNXGCGAFG 29  
|||||  
Db 840 AYCGLRPGVSSNLSAVATGNXGCGAFG 868  
|||||

RESULT 11  
US-09-511-477-6

; Sequence 6, Application US/09511477  
; Patent No. 6337202  
; GENERAL INFORMATION:  
; APPLICANT: JACOBSON, Myron K.  
; APPLICANT: JACOBSON, Elaine L.  
; APPLICANT: AME, Jean-Christophe  
; APPLICANT: LIN, Winston  
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE (PARG) EN  
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV  
; TITLE OF INVENTION: THEREWITH  
; FILE REFERENCE: NIAD 201  
; CURRENT APPLICATION NUMBER: US/09/511.477  
; CURRENT FILING DATE: 2000-02-23  
; PRIOR APPLICATION NUMBER: 09/302,812  
; PRIOR FILING DATE: 1999-04-30  
; NUMBER OF SEQ ID NOS: 38  
; SEQ ID NO 6  
; LENGTH: 968  
; TYPE: PRT  
; ORGANISM: Mus musculus  
; FEATURE:  
US-09-511-477-6

Query Match 94.7%; Score 144; DB 4; Length 968;  
Best Local Similarity 93.1%; Pred. No. 3.5e-12;  
Matches 27; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AYCGLRPGVSSENLSAVATGNXGCGAFG 29  
DB 840 AYCGLRPGVSSENLSAVATGNXGCGAFG 868

## RESULT 12

US-09-511-507-6  
; Sequence 6, Application US/09511507  
; Patent No. 6395543  
; GENERAL INFORMATION:  
; APPLICANT: JACOBSON, Myron K.  
; APPLICANT: JACOBSON, Elaine L.  
; APPLICANT: AME, Jean-Christophe  
; APPLICANT: LIN, Winston  
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE (PARG) EN  
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV  
; TITLE OF INVENTION: THEREWITH  
; FILE REFERENCE: NIAD 201  
; CURRENT APPLICATION NUMBER: US/09/511.507  
; CURRENT FILING DATE: 2000-02-23  
; PRIOR APPLICATION NUMBER: 09/302,812  
; PRIOR FILING DATE: 1999-04-30  
; NUMBER OF SEQ ID NOS: 38  
; SEQ ID NO 6  
; LENGTH: 968  
; TYPE: PRT  
; ORGANISM: Mus musculus  
; FEATURE:  
US-09-511-507-6

Query Match 94.7%; Score 144; DB 4; Length 968;  
Best Local Similarity 93.1%; Pred. No. 3.5e-12;  
Matches 27; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AYCGLRPGVSSENLSAVATGNXGCGAFG 29  
DB 840 AYCGLRPGVSSENLSAVATGNXGCGAFG 868

## RESULT 13

US-09-302-812-8  
; Sequence 8, Application US/09302812B  
; Patent No. 6333148  
; GENERAL INFORMATION:  
; APPLICANT: JACOBSON, Myron K.  
; APPLICANT: JACOBSON, Elaine L.

; APPLICANT: AM, Jean-Christophe  
; APPLICANT: LIN, Winston  
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE (PARG) ENZ  
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIVE  
; TITLE OF INVENTION: THEREWITH  
; FILE REFERENCE: NIAD 201  
; CURRENT APPLICATION NUMBER: US/09/302,812B  
; CURRENT FILING DATE: 1999-04-30  
; EARLIER APPLICATION NUMBER: 60/083,768  
; EARLIER FILING DATE: 1998-05-01  
; NUMBER OF SEQ ID NOS: 38  
; SEQ ID NO 8  
; LENGTH: 768  
; TYPE: PRT  
; ORGANISM: Drosophila melanogaster  
; FEATURE:  
US-09-302-812-8

Query Match 45.4%; Score 69; DB 4; Length 768;  
Best Local Similarity 48.6%; Pred. No. 0.12;  
Matches 17; Conservative 1; Mismatches 3; Indels 14; Gaps 2;

QY 1 AYCGLR-----PGVSSENLSAVATGNXGCGAFG 29  
DB 478 AYIGFVHMVTPPPG-----VATGNWCGGAFG 504

## RESULT 14

US-09-511-477-8  
; Sequence 8, Application US/09511477  
; Patent No. 6337202  
; GENERAL INFORMATION:  
; APPLICANT: JACOBSON, Myron K.  
; APPLICANT: JACOBSON, Elaine L.  
; APPLICANT: AME, Jean-Christophe  
; APPLICANT: LIN, Winston  
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE (PARG) EN  
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV  
; TITLE OF INVENTION: THEREWITH  
; FILE REFERENCE: NIAD 201  
; CURRENT APPLICATION NUMBER: US/09/511.477  
; CURRENT FILING DATE: 2000-02-23  
; PRIOR APPLICATION NUMBER: 09/302,812  
; PRIOR FILING DATE: 1999-04-30  
; NUMBER OF SEQ ID NOS: 38  
; SEQ ID NO 8  
; LENGTH: 768  
; TYPE: PRT  
; ORGANISM: Drosophila melanogaster  
; FEATURE:  
US-09-511-477-8

Query Match 45.4%; Score 69; DB 4; Length 768;  
Best Local Similarity 48.6%; Pred. No. 0.12;  
Matches 17; Conservative 1; Mismatches 3; Indels 14; Gaps 2;

QY 1 AYCGLR-----PGVSSENLSAVATGNXGCGAFG 29  
DB 478 AYIGFVHMVTPPPG-----VATGNWCGGAFG 504

## RESULT 15

US-09-511-507-8  
; Sequence 8, Application US/09511507  
; Patent No. 6395543  
; GENERAL INFORMATION:  
; APPLICANT: JACOBSON, Myron K.  
; APPLICANT: JACOBSON, Elaine L.  
; APPLICANT: AME, Jean-Christophe  
; APPLICANT: LIN, Winston  
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE (PARG) EN  
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV  
; TITLE OF INVENTION: THEREWITH

```

; FILE REFERENCE: N1AD 201
; CURRENT APPLICATION NUMBER: US/09/511,507
; CURRENT FILING DATE: 2000-02-23
; PRIOR APPLICATION NUMBER: 09/302,812
; PRIOR FILING DATE: 1999-04-30
; NUMBER OF SEQ ID NOS: 38
; SEQ ID NO 8
; LENGTH: 768
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
; FEATURE:
US-09-511-507-8

Query Match      45.4%; Score 69; DB 4; Length 768;
Best Local Similarity 48.8%; Pred. No. 0.12;
Matches 17; Conservative 1; Mismatches 3; Indels 14; Gaps 2;

QY      1 AYCGLR-----PGVSSNLSSAVATGNXGCGAFG 29
      |||||
      |||||
Db      478 AYIGFVHMVTPPG-----VATGNWCGGAFG 504

Search completed: May 26, 2004, 18:49:31
Job time : 1.92926 secs

```

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 26, 2004, 18:46:21 ; Search time 2.41959 Seconds

(without alignments)  
3343.018 Million cell updates/sec

Title: US-09-302-812-12

Perfect score: 152

Sequence: 1 AYCGFLRPGVSSSENLSAVATNGXGCGAFG 29

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1149313 seqs, 278921704 residues

Total number of hits satisfying chosen parameters: 1149313

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:

1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*  
2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pep.\*  
3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep.\*  
4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep.\*  
5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep.\*  
6: /cgn2\_6/ptodata/1/pubpaa/PCTUS\_PUBCOMB.pep.\*  
7: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep.\*  
8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep.\*  
9: /cgn2\_6/ptodata/1/pubpaa/US08A\_PUBCOMB.pep.\*  
10: /cgn2\_6/ptodata/1/pubpaa/US09B\_PUBCOMB.pep.\*  
11: /cgn2\_6/ptodata/1/pubpaa/US09C\_PUBCOMB.pep.\*  
12: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep.\*  
13: /cgn2\_6/ptodata/1/pubpaa/US10A\_PUBCOMB.pep.\*  
14: /cgn2\_6/ptodata/1/pubpaa/US10B\_PUBCOMB.pep.\*  
15: /cgn2\_6/ptodata/1/pubpaa/US10C\_PUBCOMB.pep.\*  
16: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep.\*  
17: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep.\*  
18: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	150	98.7	29	9	US-09-973-451-12
2	149	98.0	976	9	US-09-973-451-4
3	149	98.0	977	9	US-09-973-451-2
4	144	94.7	968	9	US-09-973-451-6
5	69	45.4	768	9	US-09-973-451-8
6	67	44.1	546	12	US-10-425-114-60000
7	66	43.4	180	12	US-10-424-599-156445
8	60.5	39.8	543	14	US-10-331-061-10
9	55.5	36.5	768	15	US-10-369-493-2214
10	53	34.9	75	9	US-09-993-844-46
11	53	34.9	75	16	US-10-633-438-32
12	53	34.9	403	14	US-10-251-385-114
13	53	34.9	403	14	US-10-251-385-224
14	53	34.9	403	14	US-10-225-567A-540
15	53	34.9	403	14	US-10-290-078-18

16	53	34.9	403	15	US-10-353-690-10	Sequence 10, Appl
17	53	34.9	426	12	US-10-311-671-1	Sequence 1, Appl
18	53	34.9	445	15	US-10-240-145-53	Sequence 53, Appl
19	53	34.9	445	15	US-10-240-145-139	Sequence 139, App
20	53	34.9	2762	12	US-10-142-426-13	Sequence 13, Appl
21	53	34.9	2762	14	US-10-123-155-13	Sequence 13, Appl
22	53	34.9	2762	14	US-10-146-731-13	Sequence 13, Appl
23	53	34.9	2762	14	US-10-140-472-13	Sequence 13, Appl
24	53	34.9	2762	14	US-10-141-761-13	Sequence 13, Appl
25	53	34.9	2762	14	US-10-142-885-13	Sequence 13, Appl
26	53	34.9	2762	14	US-10-158-790-13	Sequence 13, Appl
27	53	34.9	2762	15	US-10-137-871-13	Sequence 13, Appl
28	53	34.9	2762	15	US-10-140-923-13	Sequence 13, Appl
29	53	34.9	2762	15	US-10-141-756-13	Sequence 13, Appl
30	53	34.9	2762	15	US-10-141-759-13	Sequence 13, Appl
31	53	34.9	2762	15	US-10-140-805-13	Sequence 13, Appl
32	53	34.9	2762	15	US-10-140-864-13	Sequence 13, Appl
33	52	34.2	181	12	US-10-424-599-177700	Sequence 177700,
34	51	33.6	371	15	US-10-369-493-19533	Sequence 19533, A
35	51	33.6	379	9	US-09-738-626-5142	Sequence 5142, Ap
36	50.5	33.2	501	12	US-10-282-122A-63359	Sequence 63359, A
37	50	32.9	372	12	US-10-424-599-269980	Sequence 269980,
38	49.5	32.6	726	9	US-09-973-451-10	Sequence 10, Appl
39	49	32.2	559	14	US-10-271-697-4	Sequence 4, Appl
40	49	32.2	568	9	US-09-815-242-5140	Sequence 5140, Ap
41	49	32.2	568	12	US-10-282-122A-43547	Sequence 43547, A
42	49	32.2	575	12	US-10-425-114-45940	Sequence 45940, A
43	49	32.2	616	10	US-09-946-374-24	Sequence 24, Appl
44	49	32.2	616	12	US-10-206-915-158	Sequence 158, App
45	49	32.2	616	12	US-10-199-670-158	Sequence 158, App

#### ALIGNMENTS

#### RESULT 1

US-09-973-451-12  
; Sequence 12, Application US/09973451  
; Patent No. US20020132328A1  
; GENERAL INFORMATION:  
; APPLICANT: JACOBSON, Myron K.  
; APPLICANT: JACOBSON, Elaine L.  
; APPLICANT: AM, Jean-Christophe  
; APPLICANT: LIN, Winston  
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE  
; TITLE OF INVENTION: (PARG) ENZYMES.  
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIVE  
; TITLE OF INVENTION: THEREWITH  
; FILE REFERENCE: NIAID 201  
; CURRENT APPLICATION NUMBER: US/09/973,451  
; CURRENT FILING DATE: 2001-10-09  
; PRIOR APPLICATION NUMBER: US/09/302,812  
; PRIOR FILING DATE: 1999-04-30  
; PRIOR APPLICATION NUMBER: 60/083,768  
; PRIOR FILING DATE: 1998-05-01  
; NUMBER OF SEQ ID NOS: 38  
; SEQ ID NO 12  
; LENGTH: 29  
; TYPE: PRT  
; ORGANISM: Bos taurus  
; FEATURE:  
US-09-973-451-12

Query Match 98.7%; Score 150; DB 9; Length 29;

Best Local Similarity 100.0%; Pred. No. 1.6e-14; Indels 0; Gaps 0;  
Matches 29; Conservative 0; Mismatches 0;

QY 1 AYCGFLRPGVSSSENLSAVATNGXGCGAFG 29

Db 1 AYCGFLRPGVSSSENLSAVATNGXGCGAFG 29

#### RESULT 2

US-09-973-451-4  
 ; Sequence 4, Application US/09973451  
 ; Patent No. US20020132328A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: JACOBSON, Myron K.  
 ; APPLICANT: JACOBSON, Elaine L.  
 ; APPLICANT: AM, Jean-Christophe  
 ; APPLICANT: LIN, Winston  
 ; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE  
 ; TITLE OF INVENTION: (PARG) ENZYMES,  
 ; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV  
 ; TITLE OF INVENTION: THEREWITH  
 ; FILE REFERENCE: NIAD 201  
 ; CURRENT APPLICATION NUMBER: US/09/973,451  
 ; CURRENT FILING DATE: 2001-10-09  
 ; PRIOR APPLICATION NUMBER: US/09/302,812  
 ; PRIOR FILING DATE: 1999-04-30  
 ; PRIOR APPLICATION NUMBER: 60/083,768  
 ; PRIOR FILING DATE: 1998-05-01  
 ; NUMBER OF SEQ ID NOS: 38  
 ; SEQ ID NO 4  
 ; LENGTH: 976  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 ; FEATURE:  
 ;  
 US-09-973-451-4

Query Match 98.0%; Score 149; DB 9; Length 976;  
 Best Local Similarity 96.6%; Pred. No. 1.1e-12;  
 Matches 28; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AYCGLRPGVSSSENLSAVATGNKCGAGF 29  
 DB 848 AYCGLRPGVSSSENLSAVATGNWCGAGF 876

RESULT 3  
 US-09-973-451-2  
 ; Sequence 2, Application US/09973451  
 ; Patent No. US20020132328A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: JACOBSON, Myron K.  
 ; APPLICANT: JACOBSON, Elaine L.  
 ; APPLICANT: AM, Jean-Christophe  
 ; APPLICANT: LIN, Winston  
 ; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE  
 ; TITLE OF INVENTION: (PARG) ENZYMES,  
 ; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV  
 ; TITLE OF INVENTION: THEREWITH  
 ; FILE REFERENCE: NIAD 201  
 ; CURRENT APPLICATION NUMBER: US/09/973,451  
 ; CURRENT FILING DATE: 2001-10-09  
 ; PRIOR APPLICATION NUMBER: US/09/302,812  
 ; PRIOR FILING DATE: 1999-04-30  
 ; PRIOR APPLICATION NUMBER: 60/083,768  
 ; PRIOR FILING DATE: 1998-05-01  
 ; NUMBER OF SEQ ID NOS: 38  
 ; SEQ ID NO 2  
 ; LENGTH: 977  
 ; TYPE: PRT  
 ; ORGANISM: Bos taurus  
 ; FEATURE:  
 ;  
 US-09-973-451-2

Query Match 98.0%; Score 149; DB 9; Length 977;  
 Best Local Similarity 96.6%; Pred. No. 1.1e-12;  
 Matches 28; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AYCGLRPGVSSSENLSAVATGNKCGAGF 29  
 DB 849 AYCGLRPGVSSSENLSAVATGNWCGAGF 877

RESULT 4  
 US-09-973-451-6  
 ; Sequence 6, Application US/09973451  
 ; Patent No. US20020132328A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: JACOBSON, Myron K.  
 ; APPLICANT: JACOBSON, Elaine L.  
 ; APPLICANT: AM, Jean-Christophe  
 ; APPLICANT: LIN, Winston  
 ; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE  
 ; TITLE OF INVENTION: (PARG) ENZYMES,  
 ; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV  
 ; TITLE OF INVENTION: THEREWITH  
 ; FILE REFERENCE: NIAD 201  
 ; CURRENT APPLICATION NUMBER: US/09/973,451  
 ; CURRENT FILING DATE: 2001-10-09  
 ; PRIOR APPLICATION NUMBER: US/09/302,812  
 ; PRIOR FILING DATE: 1999-04-30  
 ; PRIOR APPLICATION NUMBER: 60/083,768  
 ; PRIOR FILING DATE: 1998-05-01  
 ; NUMBER OF SEQ ID NOS: 38  
 ; SEQ ID NO 6  
 ; LENGTH: 968  
 ; TYPE: PRT  
 ; ORGANISM: Mus musculus  
 ; FEATURE:  
 ;  
 US-09-973-451-6

Query Match 94.7%; Score 144; DB 9; Length 968;  
 Best Local Similarity 93.1%; Pred. No. 6e-12;  
 Matches 27; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AYCGLRPGVSSSENLSAVATGNKCGAGF 29  
 DB 840 AYCGLRPGVSSSENLSAVATGNWCGAGF 868

RESULT 5  
 US-09-973-451-8  
 ; Sequence 8, Application US/09973451  
 ; Patent No. US20020132328A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: JACOBSON, Myron K.  
 ; APPLICANT: JACOBSON, Elaine L.  
 ; APPLICANT: AM, Jean-Christophe  
 ; APPLICANT: LIN, Winston  
 ; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE  
 ; TITLE OF INVENTION: (PARG) ENZYMES,  
 ; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV  
 ; TITLE OF INVENTION: THEREWITH  
 ; FILE REFERENCE: NIAD 201  
 ; CURRENT APPLICATION NUMBER: US/09/973,451  
 ; CURRENT FILING DATE: 2001-10-09  
 ; PRIOR APPLICATION NUMBER: US/09/302,812  
 ; PRIOR FILING DATE: 1999-04-30  
 ; PRIOR APPLICATION NUMBER: 60/083,768  
 ; PRIOR FILING DATE: 1998-05-01  
 ; NUMBER OF SEQ ID NOS: 38  
 ; SEQ ID NO 8  
 ; LENGTH: 768  
 ; TYPE: PRT  
 ; ORGANISM: Drosophila melanogaster  
 ; FEATURE:  
 ;  
 US-09-973-451-8

Query Match 45.4%; Score 69; DB 9; Length 768;  
 Best Local Similarity 48.6%; Pred. No. 0.31;  
 Matches 17; Conservative 1; Mismatches 3; Indels 14; Gaps 2;

QY 1 AYCGLR-----PGVSSSENLSAVATGNKCGAGF 29  
 DB 478 AYIGFVHMVTPPPG-----VATGNWCGAGF 504

```
RESULT 6
US-10-425-114-60000
; Sequence 60000, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; PRIOR FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 60000
; LENGTH: 546
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB3689-262-E12_FLI.pep
US-10-425-114-60000
Query Match 44.1%; Score 67; DB 12; Length 546;
Best Local Similarity 30.6%; Pred. No. 0.42; Indels 20; Gaps 1;
Matches 15; Conservative 4; Mismatches 4;
QY 1 AYCGLR-----PGVSSNLSAVATGNKCGGAFG 29
Db 407 AFGCFDQSKHLYAKLFQDLHNKDPSSINSSEYGVSTGNWCGGAFG 455

RESULT 7
US-10-424-599-156445
; Sequence 156445, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; PRIOR FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 156445
; LENGTH: 180
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_112291C.1.pep
US-10-424-599-156445
Query Match 43.4%; Score 66; DB 12; Length 180;
Best Local Similarity 61.9%; Pred. No. 0.17; Indels 7; Gaps 0;
Matches 13; Conservative 1; Mismatches 7;
QY 9 GVSSENLSAVATGNKCGGAFG 29
Db 67 GMDQGNIGVATGNWCGGAFG 87

RESULT 8
US-10-331-061-10
; Sequence 10, Application US/10331061
; Publication No. US20030101486A1
; GENERAL INFORMATION:
; APPLICANT: Lasener, Michael
; APPLICANT: Metz, James G
; APPLICANT: Facciotti, Daniel
; TITLE OF INVENTION: SCHIZOCHYTRIUM PKS GENES
; FILE REFERENCE: CGNE.131.02US
; CURRENT APPLICATION NUMBER: US/10/331,061
; CURRENT FILING DATE: 2002-12-27
; PRIOR APPLICATION NUMBER: US/09/231,899
; PRIOR FILING DATE: 1999-01-14
; PRIOR APPLICATION NUMBER: 60/048,650
; PRIOR FILING DATE: 1997-06-04
; PRIOR APPLICATION NUMBER: 09/090,793
; PRIOR FILING DATE: 1998-06-04
; NUMBER OF SEQ ID NOS: 86
; SOFTWARE: Patent in Ver. 2.0
; SEQ ID NO 10
; LENGTH: 543
; TYPE: PRT
; ORGANISM: Shewanella putrefaciens
US-10-331-061-10
Query Match 39.8%; Score 60.5; DB 14; Length 543;
Best Local Similarity 46.7%; Pred. No. 3.6; Indels 3; Gaps 2;
Matches 14; Conservative 4; Mismatches 9;
QY 2 YCGFLRPGVSSNLSAVATGNK--CGAFG 29
Db 99 YAGAMANGISSEEL-VIALQGAGILCGSFG 127

RESULT 9
US-10-369-493-2214
; Sequence 2214, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 2214
; LENGTH: 768
; TYPE: PRT
; ORGANISM: Schizosaccharomyces pombe
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(768)
; OTHER INFORMATION: unsure at all Xaa locations
US-10-369-493-2214
Query Match 36.5%; Score 55.5; DB 15; Length 768;
Best Local Similarity 53.8%; Pred. No. 28; Indels 3; Gaps 1;
Matches 14; Conservative 1; Mismatches 8;
QY 1 AYCGLRPGVSSNLSAVATGNKCG 26
Db 498 ALCDF---GLSKANLSANATTTFCG 520

RESULT 10
US-09-993-844-46
; Sequence 46, Application US/09993844
; Patent No. US20020106739A1
; GENERAL INFORMATION:
; APPLICANT: Oakley, Robert H.
; APPLICANT: Barak, Lawrence S.
```



APPLICANT: Laporte, Stephane A.  
APPLICANT: Caron, Marc G.  
TITLE OF INVENTION: Modified G-Protein Coupled Receptors  
FILE REFERENCE: 033072-026  
CURRENT APPLICATION NUMBER: US/09/993,844  
CURRENT FILING DATE: 2001-11-05  
PRIOR APPLICATION NUMBER: US 60/245,772  
PRIOR FILING DATE: 2000-11-03  
PRIOR APPLICATION NUMBER: US 60/260,363  
PRIOR FILING DATE: 2001-01-08  
NUMBER OF SEQ ID NOS: 81  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 46  
LENGTH: 75  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-993-844-46

Query Match 34.9%; Score 53; DB 9; Length 75;  
Best Local Similarity 45.2%; Pred. No. 4.8;  
Matches 14; Conservative 4; Mismatches 11; Indels 2; Gaps 1;

QY 1 AYCGLRPGVSSSENLSAVATGNKGC--GAFG 29  
Db 24 ACCHRLRPRHSSHSLSRMTTGSTLCDVGSGLG 54

RESULT 11  
US-10-633-438-32  
Sequence 32, Application US/10633438  
Publication No. US20040091946A1  
GENERAL INFORMATION:  
APPLICANT: Oakley, Robert H.  
APPLICANT: Barak, Lawrence S.  
APPLICANT: Laporte, Stephane A.  
APPLICANT: Caron, Marc G.  
TITLE OF INVENTION: Methods of Screening Compositions for G Protein-Coupled Receptor  
TITLE OF INVENTION: Desensitization Inhibitory Activity  
FILE REFERENCE: 033072-044  
CURRENT APPLICATION NUMBER: US/10/633,438  
CURRENT FILING DATE: 2003-08-01  
PRIOR APPLICATION NUMBER: US 09/993,844  
PRIOR FILING DATE: 2001-11-05  
PRIOR APPLICATION NUMBER: US 60/245,772  
PRIOR FILING DATE: 2000-11-03  
PRIOR APPLICATION NUMBER: US 60/260,363  
PRIOR FILING DATE: 2001-01-08  
NUMBER OF SEQ ID NOS: 65  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 32  
LENGTH: 75  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-633-438-32

Query Match 34.9%; Score 53; DB 16; Length 75;  
Best Local Similarity 45.2%; Pred. No. 4.8;  
Matches 14; Conservative 4; Mismatches 11; Indels 2; Gaps 1;

QY 1 AYCGLRPGVSSSENLSAVATGNKGC--GAFG 29  
Db 24 ACCHRLRPRHSSHSLSRMTTGSTLCDVGSGLG 54

RESULT 12  
US-10-251-385-114  
Sequence 114, Application US/10251385  
Publication No. US20030105292A1  
GENERAL INFORMATION:  
APPLICANT: Behan, Dominic P.  
APPLICANT: Chalmers, Derek T.  
APPLICANT: Liaw, Chen W.  
TITLE OF INVENTION: No. US20030105292A1-Endogenous, Constitutively Activated Human G

TITLE OF INVENTION: Protein-Coupled  
TITLE OF INVENTION: Receptors  
FILE REFERENCE: AREN-0040  
CURRENT APPLICATION NUMBER: US/10/251,385  
CURRENT FILING DATE: 2002-09-20  
PRIOR APPLICATION NUMBER: US/09/170,496  
PRIOR FILING DATE: 1998-10-13  
NUMBER OF SEQ ID NOS: 294  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 114  
LENGTH: 403  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-251-385-114

Query Match 34.9%; Score 53; DB 14; Length 403;  
Best Local Similarity 45.2%; Pred. No. 31;  
Matches 14; Conservative 4; Mismatches 11; Indels 2; Gaps 1;

QY 1 AYCGLRPGVSSSENLSAVATGNKGC--GAFG 29  
Db 352 ACCHRLRPRHSSHSLSRMTTGSTLCDVGSGLG 382

RESULT 13  
US-10-251-385-224  
Sequence 224, Application US/10251385  
Publication No. US20030105292A1  
GENERAL INFORMATION:  
APPLICANT: Behan, Dominic P.  
APPLICANT: Chalmers, Derek T.  
APPLICANT: Liaw, Chen W.  
TITLE OF INVENTION: No. US20030105292A1-Endogenous, Constitutively Activated Human G  
TITLE OF INVENTION: Protein-Coupled  
TITLE OF INVENTION: Receptors  
FILE REFERENCE: AREN-0040  
CURRENT APPLICATION NUMBER: US/10/251,385  
CURRENT FILING DATE: 2002-09-20  
PRIOR APPLICATION NUMBER: US/09/170,496  
PRIOR FILING DATE: 1998-10-13  
NUMBER OF SEQ ID NOS: 294  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 224  
LENGTH: 403  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-251-385-224

Query Match 34.9%; Score 53; DB 14; Length 403;  
Best Local Similarity 45.2%; Pred. No. 31;  
Matches 14; Conservative 4; Mismatches 11; Indels 2; Gaps 1;

QY 1 AYCGLRPGVSSSENLSAVATGNKGC--GAFG 29  
Db 352 ACCHRLRPRHSSHSLSRMTTGSTLCDVGSGLG 382

RESULT 14  
US-10-225-567A-540  
Sequence 540, Application US/10225567A  
Publication No. US20030113798A1  
GENERAL INFORMATION:  
APPLICANT: Lifespan Biosciences  
APPLICANT: Brown, Joseph P.  
APPLICANT: Burmer, Glenna C.  
APPLICANT: Roush, Christine L.  
TITLE OF INVENTION: ANTIGENIC PEPTIDES AND ANTIBODIES FOR G PROTEIN-COUPLED RECEPTORS  
FILE REFERENCE: 1920-4-4  
CURRENT APPLICATION NUMBER: US/10/225,567A  
CURRENT FILING DATE: 2001-12-19  
PRIOR APPLICATION NUMBER: 60/257,144  
PRIOR FILING DATE: 2000-12-19  
NUMBER OF SEQ ID NOS: 2292

```
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 540
; LENGTH: 403
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-225-567A-540

Query Match      34.9%; Score 53; DB 14; Length 403;
Best Local Similarity 45.2%; Pred.No. 31;
Matches 14; Conservative 4; Mismatches 11; Indels 2; Gaps 1;

QY      1 AYCGLRPGVSSNLSAVATGNXGC--GAFG 29
Db      352 ACCHRLRPRHSHSLSRMTTGSTLCDVGSIG 382

RESULT 15
US-10-290-078-18
; Sequence 18, Application US/10290078
; Publication No. US20030124596A1
; GENERAL INFORMATION:
; APPLICANT: Carroll, Joseph A.
; TITLE OF INVENTION: Methods and Compositions for Treating
; TITLE OF INVENTION: Hematological Disorders Using 232, 2059, 10630, 12848, 13875,
; TITLE OF INVENTION: 14395, 14618, 17692 or 58874
; FILE REFERENCE: MPI2001-288P1(M)
; CURRENT APPLICATION NUMBER: US/10/290,078
; CURRENT FILING DATE: 2002-11-07
; NUMBER OF SEQ ID NOS: 27
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 18
; LENGTH: 403
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-290-078-18

Query Match      34.9%; Score 53; DB 14; Length 403;
Best Local Similarity 45.2%; Pred.No. 31;
Matches 14; Conservative 4; Mismatches 11; Indels 2; Gaps 1;

QY      1 AYCGLRPGVSSNLSAVATGNXGC--GAFG 29
Db      352 ACCHRLRPRHSHSLSRMTTGSTLCDVGSIG 382

Search completed: May 26, 2004, 19:19:01
Job time : 3.41959 secs
```





Query Match 32.2%; Score 49; DB 2; Length 568;  
 Best Local Similarity 40.7%; Pred. No. 39;  
 Matches 11; Conservative 3; Mismatches 9; Indels 4; Gaps 1;

QY 3 CGFLRPGVSSSENLSAVATGNKCGGAFG 29  
 ||:| :|:| |||||  
 276 CGYL----DYQGVIAIAQFACGGAFG 298  
 ||:| :|:| |||||

RESULT 10  
 D82279 sodium/alanine symporter VC0784 [imported] - Vibrio cholerae (strain N16961 serogroup O1)  
 C:Species: Vibrio cholerae  
 C>Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 02-Feb-2001  
 C:Accession: D82279  
 R.Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.J.;  
 Richardson, D.; Ermlaeva, M.D.; Vamachavan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellers, P.  
 1, L.R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.  
 Nature 406, 477-483, 2000  
 A:Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.  
 A:Reference number: A82035; MUID:20406833; PMID:10952301  
 A:Accession: D82279  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-516 <HEI>  
 A:Cross-references: GB:AE004164; GB:AE003852; NID:g9655232; PIDN:AAF93949.1; GSPDB:GN001  
 A:Experimental source: serogroup O1; strain N16961; biotype El Tor  
 C:Genetics:  
 A:Gene: VC0784  
 A:Map position: 1  
 C:Superfamily: sodium-dependent D-alanine/glycine transport protein

Query Match 31.9%; Score 48.5; DB 2; Length 516;  
 Best Local Similarity 44.4%; Pred. No. 42;  
 Matches 12; Conservative 2; Mismatches 8; Indels 5; Gaps 1;

QY 3 CGFLRPGVSSSENLSAVATGNKCGGAFG 29  
 ||:| ||||| :|:| |||||  
 152 CGVLLPFGVQNSI----GNAVESAFG 173

RESULT 11  
 AE2658 glycolate oxidase iron-sulfur subunit gicF [imported] - Agrobacterium tumefaciens (strain  
 C:Species: Agrobacterium tumefaciens  
 C>Date: 11-Jan-2002 #sequence\_revision 11-Jan-2002 #text\_change 18-Aug-2003  
 C:Accession: AE2658  
 R.Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, I.;  
 erage, G.; Gallet, W.; Grant, C.; Guenther, D.; Kutayavin, T.; Levy, R.; Li, M.; McClell  
 ; Karp, P.; Romero, P.; Zhang, S.  
 Science 294, 2317-2323, 2001  
 A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,  
 ster, E.W.  
 A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.  
 A:Reference number: AB2577; MUID:21608550; PMID:11743193  
 A:Accession: AE2658  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-433 <KUR>  
 A:Cross-references: GB:AE008688; PIDN:AAL41683.1; PID:g17739028; GSPDB:GN00186  
 A:Experimental source: strain C58 (Dupont)  
 C:Genetics:  
 A:Gene: gicF  
 A:Map position: circular chromosome  
 C:Superfamily: (S)-2-hydroxy-acid oxidase

Query Match 31.6%; Score 48; DB 2; Length 433;  
 Best Local Similarity 34.1%; Pred. No. 41;  
 Matches 14; Conservative 5; Mismatches 4; Indels 18; Gaps 3;

QY 3 CG-----FLRPGVSS-----ENLSA-----VATGNKGC 25  
 ||:| :|:| ||||| :|:| :|:| |||||

R:Du, Z.  
submitted to the EMBL Data Library, July 1995  
A:Description: sequence of C. elegans cosmid C30B5.  
A:Reference number: S59417  
A:Accession: T15705  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-411 <DUZ>  
A:Cross-references: EMBL:U23450; NID:g733552; PID:g733553; PIDN:AAC46739.1; CESP:C30B5.3  
A:Experimental source: strain Bristol N2  
C:Genetics:  
A:Gene: CESP:C30B5.3  
A:Introns: 12/2; 91/1; 349/3

Query Match 31.2%; Score 47.5; DB 2; Length 411;  
Best Local Similarity 45.8%; Pred. No. 46;  
Matches 11; Conservative 3; Mismatches 5; Indels 5; Gaps 1;

QY 4 GFLRPGVSSSENLSAVATGNKCGA 27  
||| ||| :| ||| ||| :  
Db 177 GFLNPG-----MAQVARGNLGFGS 195

RESULT 15  
AB2501  
hypothetical protein all7186 [imported] - Nostoc sp. (strain PCC 7120) plasmid pCC7120a1  
C:Species: Nostoc sp. PCC 7120  
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120  
C:Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 09-Dec-2002  
C:Accession: AB2501  
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi  
R:Kaneko, T.; Nakamura, Y.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Tabata, S.  
DNA Res. 8, 205-213, 2001  
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena  
A:Reference number: AB1807; MUID:21595285; PMID:11759840  
A:Accession: AB2501  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-129 <KUR>  
A:Cross-references: GB:BA000020; PIDN:BA078270.1; PID:g17135724; GSPDB:GN00180  
A:Experimental source: strain PCC 7120  
C:Genetics:  
A:Gene: all7186  
A:Genome: plasmid

Query Match 30.9%; Score 47; DB 2; Length 129;  
Best Local Similarity 56.2%; Pred. No. 17;  
Matches 9; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 5 FLRPGVSSSENLSAVAT 20  
||| ||| ||| ||| :|  
Db 30 FARPAQVQENLTAET 45

Search completed: May 26, 2004, 18:47:59  
Job time : 2.79776 secs

Db 360 CGSAGTYNLOPEISSKLKARKVNIETRADVIATGNIGC 400

RESULT 12  
B97440  
glycolate oxidase iron-sulfur chain [imported] - Agrobacterium tumefaciens (strain C58,  
C:Species: Agrobacterium tumefaciens  
C:Date: 30-Sep-2001 #sequence\_revision 30-Sep-2001 #text\_change 18-Aug-2003  
C:Accession: B97440  
R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman,  
A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.;  
Science 294, 2323-2328, 2001  
A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tumefaciens  
A:Reference number: A97359; MUID:21608551; PMID:11743194  
A:Accession: B97440  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-433 <KUR>  
A:Cross-references: GB:AE007869; PIDN:AAK86475.1; PID:g15155623; GSPDB:GN00169  
C:Genetics:  
A:Gene: AGR\_C1194  
A:Map position: circular chromosome  
C:Superfamily: (S)-2-hydroxy-acid oxidase

Query Match 31.6%; Score 48; DB 2; Length 433;  
Best Local Similarity 34.1%; Pred. No. 41;  
Matches 14; Conservative 5; Mismatches 4; Indels 18; Gaps 3;

QY 3 CG-----FLRPGVSS-----ENLSA-----VATGNKGC 25  
||| ||| :| ||| ||| :  
Db 360 CGSAGTYNLOPEISSKLKARKVNIETRADVIATGNIGC 400

RESULT 13  
S77484  
preprotein translocase secY - Synechocystis sp. (strain PCC 6803)  
N:Alternate names: preprotein translocase secY chain; protein sll1814  
C:Species: Synechocystis sp.  
A:Variety: PCC 6803  
C:Date: 25-Apr-1997 #sequence\_revision 25-Apr-1997 #text\_change 20-Jun-2000  
C:Accession: S77484  
R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.;  
O. K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda  
DNA Res. 3, 109-136, 1996  
A:Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocystis  
A:Reference number: S74322; MUID:97061201; PMID:8905231  
A:Accession: S77484  
A:Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-442 <KAN>  
A:Cross-references: EMBL:D90905; GB:AB001339; NID:g1652360; PIDN:BAAL17331.1; PID:g165240  
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996  
C:Genetics:  
A:Gene: secY  
C:Superfamily: preprotein translocase secY  
C:Keywords: protein transport; transmembrane protein

Query Match 31.6%; Score 48; DB 2; Length 442;  
Best Local Similarity 56.2%; Pred. No. 42;  
Matches 9; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 14 NLSAVATGNKCGAFG 29  
||| ||| ||| ||| :|  
Db 289 SLAFATGNEGLGFGF 304

RESULT 14  
T15705  
hypothetical protein C30B5.3 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 23-Mar-2001  
C:Accession: T15705

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 26, 2004, 17:20:49 ; Search time 0.464631 Seconds  
(without alignments)  
3249.964 Million cell updates/sec

Title: US-09-302-812-12

Perfect score: 152

Sequence: 1 AYCGFLRPGVSSNL SAVATGNKGGAGF 29

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_42.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	55.5	36.5	696	1	SCK1 SCHPO
2	51	33.6	511	1	GUNB PSEFL
3	48.5	31.9	886	1	SM6B MOUSE
4	48.5	31.9	887	1	SM6B RAT
5	48	31.6	442	1	SECV SYNY3
6	47.5	31.2	175	1	NU6M DUGDU
7	47	30.9	595	1	THD1 LYCES
8	46.5	30.6	307	1	IMP3 HUMAN
9	46.5	30.6	413	1	MAAL CLOTT
10	46	30.3	262	1	Y877 MYCTU
11	46	30.3	273	1	GYF0 YEAST
12	46	30.3	1305	1	CYAS XENLA
13	46	30.3	2358	1	YEEU ECOLI
14	46	30.3	2556	1	NTCI HUMAN
15	45.5	29.9	424	1	PYRX PSEPU
16	45	29.6	330	1	TLX1 HUMAN
17	45	29.6	359	1	THD1 SOLTU
18	45	29.6	474	1	RHOF ECOLI
19	44	28.9	211	1	SHUI HUMAN
20	44	28.9	331	1	PME ASPTU
21	44	28.9	338	1	RTCA ECOLI
22	44	28.9	339	1	RTCA SALTY
23	44	28.9	341	1	RTCA PREAE
24	44	28.9	342	1	RTCA ECO57
25	44	28.9	358	1	DPNF ORYSA
26	44	28.9	360	1	FADH AMYME
27	44	28.9	481	1	LEU2 COREF
28	44	28.9	481	1	LEU2 CORGL
29	44	28.9	485	1	LEU2 ACTTI
30	44	28.9	614	1	ASNO_BACSU
31	44	28.9	735	1	AD02 CAVPO
32	44	28.9	747	1	SPD1 NEFCL
33	44	28.9	971	1	AREA_GIBFU

```

34 44 28.9 1093 1 AF17 HUMAN
35 44 28.9 1365 1 SU22 DROME
36 44 28.9 1801 1 LMB2 RAT
37 43.5 28.6 260 1 GLO2 MOUSE
38 43.5 28.6 384 1 AGAS ECOLI
39 43.5 28.6 408 1 FXDL HUMAN
40 43.5 28.6 515 1 RPB2 METTW
41 43.5 28.6 842 1 VGLH HSBVC
42 43.5 28.6 1115 1 TBC2 CHIRE
43 43 28.3 115 1 PT22 STYPL
44 43 28.3 136 1 SC3 SCHCO
45

RESULT 1
SCK1 SCHPO STANDARD; PRT; 696 AA.
AC P50530; O9UTEF3;
DT 01-OCT-1996 (Rel. 34, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Serine/threonine-protein kinase sck1 (EC 2.7.1.37).
GN SCK1 OR SPAC1B9.02C.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96120227; PubMed=7493728;
RA Jin M., Fujita M., Culley B., Apolinario E., Yamamoto M.,
RA Maundrell K., Hoffman C.;
RT "sck1, a high copy number suppressor of defects in the CAMP-dependent
RT protein kinase pathway in fission yeast, encodes a protein homologous
RT to the Saccharomyces cerevisiae SCH9 kinase.";
RL Genetics 140:457-467 (1995).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=972;
RA MEDLINE=21848401; PubMed=11859360;
RA Wood V., Gilliam R., Rajadream M.A., Lyne M., Lyne R., Stewart A.,
RA Sgouras J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA Oliver K., O'Neil S., Pearson D., Quail M.A., Rabinowitsch E.,
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA Skelton J., Simmonds M., Squares R., Stevens S., Stevens K.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,
RA Weltjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA Gabel C., Fuchs M., Fritzc C., Holzer E., Moestl D., Hilbert H.,
RA Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,
RA Eger P., Zimmermann W., Wedler H., Wambutt R., Furnelle B.,
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forburg S.L.,
RA Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA Shpakovski G.V., Ussery D., Barrrell B.G., Nurse P.;
RT "The genome sequence of Schizosaccharomyces pombe.";
RL Nature 415:871-880 (2002)
CC -!- CATALYTIC ACTIVITY: ATP + a protein = ADP + a phosphoprotein.
CC -!- SIMILARITY: Belongs to the Ser/Thr family of protein kinases. CAMP
CC subfamily.

```

#### ALIGNMENTS

-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----

EMBL; D38108; BAA07286.1; -;  
EMBL; AL109951; CAB53053.1; -;  
PIR; S55694; S55694.  
HSSP; P05132; ICIP.  
GenDB Spombe; SPAC1B9.02c; -;  
InterPro; IPR000008; C2.  
InterPro; IPR008973; C2 CalB.  
InterPro; IPR000961; Pkinase C.  
InterPro; IPR000719; Prot kinase.  
InterPro; IPR008271; Ser\_thr\_pkin\_AS.  
InterPro; IPR002290; Ser\_thr\_pkinase.  
InterPro; IPR001245; Tyr\_pkinase.  
Pfam; PF00168; C2; 1.  
Pfam; PF00069; pkinase; 1.  
Pfam; PF00433; Pkinase C; 1.  
PRINTS; PR00109; TYRKINASE.  
ProDom; PD000001; Prot\_kinase; 1.  
SMART; SM00239; C2; 1.  
SMART; SM00133; S\_TK\_X; 1.  
SMART; SM00220; S\_TK; 1.  
PROSITE; PS00107; PROTEIN KINASE ATP; 1.  
PROSITE; PS00108; PROTEIN KINASE ST; 1.  
PROSITE; PS00011; PROTEIN KINASE DOM; 1.  
Transferase; Serine/threonine-protein kinase; ATP-binding; cAMP.  
FT DOMAIN 302 563 PROTEIN KINASE.  
FT BIND 308 316 ATP (BY SIMILARITY).  
FT BINDING 331 331 ATP (BY SIMILARITY).  
FT ACT SITE 428 428 BY SIMILARITY.  
FT CONFLICT 199 199 A -> R (IN REF. 1).  
SQ SEQUENCE 696 AA; 78594 MW; A7B05F5EB4D42AF7 CRC64;

Query Match 36.5%; Score 55.5; DB 1; Length 696;  
Best Local Similarity 53.8%; Pred. No. 1.8;  
Matches 14; Conservative 1; Mismatches 8; Indels 3; Gaps 1;  
QY 1 AYCGLRPGVSSNLSAVATGNXGCG 26  
Db 443 ALCDF--GLSKANLSANATNTFCG 465

RESULT 2  
ID GUNB\_PSEFL STANDARD; PRT; 511 AA.  
AC P18126;  
DT 01-NOV-1990 (Rel. 16, Created)  
DT 01-NOV-1990 (Rel. 16, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Endoglucanase B precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase) (Cellulase) (EGB).  
GN CELB.  
OS Pseudomonas fluorescens.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;  
OC Pseudomonadaceae; Pseudomonas.  
OX NCBI\_TaxID=294;  
RN [1]  
RP SEQUENCE FROM N.A., AND SEQUENCE OF 30-48.  
RC STRAIN-Sp Cellulosa;  
RX MEDLINE=90355836; PubMed=2117693;  
RA Gilbert H.J., Hall J., Hazlewood G.P., Ferreira L.M.A.;  
RT "The N-terminal region of an endoglucanase from Pseudomonas fluorescens subsp. cellulosa constitutes a cellulose-binding domain that is distinct from the catalytic centre."  
RT Mol. Microbiol. 4:759-767(1990).  
RL M.L. FUNCTION: THIS ENZYME CATALYZES THE ENDOHYDROLYSIS OF 1,4-BETA-

GLUCOSIDIC LINKAGES IN CELLULOSE, LICHENIN AND CEREAL BETA-D-GLUCANS. EGB IS MOST ACTIVE AGAINST BARLEY BETA-GLUCAN, BUT SHOWED SIGNIFICANT ACTIVITY AGAINST AMORPHOUS AND CRISTALLINE CELLULOSE.  
-!- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic linkages in cellulose, lichenin and cereal beta-D-glucans.  
-!- SUBCELLULAR LOCATION: Periplasmic.  
-!- SIMILARITY: Contains 1 bacterial-type cellulose-binding (CBD) domain.  
-!- SIMILARITY: Belongs to cellulase family K (family 45 of glycosyl hydrolases).

-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----

EMBL; X52615; CRA36844.1; -;  
PIR; S10527; S10527.  
HSSP; P43316; 2ENG.  
InterPro; IPR001919; Bac\_celose-bind.  
InterPro; IPR009009; Barwin\_like.  
InterPro; IPR009031; CBDX.  
InterPro; IPR008965; Cellul bind.  
InterPro; IPR002883; Dockerin CBD 5.  
InterPro; IPR000334; Glyco\_hydro\_45.  
Pfam; PF02013; CBM\_10; 1.  
Pfam; PF00553; CBM\_2; 1.  
Pfam; PF02015; Glyco\_hydro\_45; 1.  
SMART; SM00637; CBD\_II; 1.  
PROSITE; PS01140; GLYCOSYL HYDROL F45; 1.  
PROSITE; PS01140; CBD BACTERIAL; 1.  
Cellulose degradation: Hydrolase; Glycosidase; Signal; Periplasmic.  
KW SIGNAL 1 29  
FT CHAIN 30 511 ENDOGLUCANASE B.  
FT DOMAIN 30 131 CELLULOSE-BINDING (BY SIMILARITY).  
FT DOMAIN 132 173 SER-RICH (LINKER).  
FT DOMAIN 223 259 SER-RICH.  
FT DISULFID 32 127 BY SIMILARITY.  
FT ACT SITE 276 276 NUCLEOPHILE (BY SIMILARITY).  
FT ACT SITE 393 393 PROTON DONOR (BY SIMILARITY).  
SQ SEQUENCE 511 AA; 52078 MW; 3C3119D998291D8E CRC64;

Query Match 33.6%; Score 51; DB 1; Length 511;  
Best Local Similarity 45.5%; Pred. No. 6.1;  
Matches 10; Conservative 3; Mismatches 9; Indels 0; Gaps 0;

QY 8 PGVSSNLSAVATGNXGCGAFG 29  
Db 164 PGTSSSSSSSVLTGAQACNWYG 185

RESULT 3  
ID SM6B\_MOUSE STANDARD; PRT; 886 AA.  
AC O54951;  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Semaphorin 6B precursor (Semaphorin VIB) (Sema VIB) (Semaphorin N) (Sema N).  
GN SEMA6B OR SEMAN.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=98027184; PubMed=9361278;  
RA Eckhardt F., Behar O., Calautti E., Yonezawa K., Nishimoto I., Fishman M.C.;



"A novel transmembrane semaphorin can bind c-src.";  
 Mol. Cell. Neurosci. 9:409-419(1997).  
 - SUBUNIT: HOMODIMER. BINDS SPECIFICALLY THE SH3 DOMAIN OF THE  
 PROTOONCOGENE C-SRC.  
 - SUBCELLULAR LOCATION: Type I membrane protein.  
 - TISSUE SPECIFICITY: During development it is expressed in  
 subregions of the nervous system and is particularly prominent in  
 muscle. In adulthood, it is expressed ubiquitously.  
 - SIMILARITY: Belongs to the semaphorin family.  
 - SIMILARITY: Contains 1 Sema domain.  
 This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 EMBL; AF036585; AAC00493.1; -  
 MGD; MG1:1202889; Sema6b.  
 InterPro; IPR003659; Plexin-like.  
 InterPro; IPR001627; Sema.  
 Pfam; PF01403; Sema; 1.  
 SMART; SM00423; PSI; 1.  
 SMART; SM00630; Sema; 1.  
 Signal; Transmembrane; Multigene family; Neurogenesis; Glycoprotein;  
 Developmental protein.  
 SIGNAL 1 26 POTENTIAL.  
 CHAIN 27 886 SEMAPHORIN 6B.  
 DOMAIN 27 605 EXTRACELLULAR (POTENTIAL).  
 TRANSMEM 606 626 POTENTIAL.  
 DOMAIN 627 886 CYTOPLASMIC (POTENTIAL).  
 DOMAIN 239 549 SEMA.  
 DOMAIN 751 754 POLY-LEU.  
 CARBOHYD 75 75 N-LINKED (GLCNAC. .) (POTENTIAL).  
 CARBOHYD 156 156 N-LINKED (GLCNAC. .) (POTENTIAL).  
 CARBOHYD 292 292 N-LINKED (GLCNAC. .) (POTENTIAL).  
 CARBOHYD 387 387 N-LINKED (GLCNAC. .) (POTENTIAL).  
 CARBOHYD 442 442 N-LINKED (GLCNAC. .) (POTENTIAL).  
 CARBOHYD 463 463 N-LINKED (GLCNAC. .) (POTENTIAL).  
 SEQUENCE 886 AA; 95466 MW; 55F56D125CDA574D CRC64;  
 Query Match 31.9%; Score 48.5; DB 1; Length 886;  
 Best Local Similarity 35.3%; Pred. No. 26;  
 Matches 12; Conservative 5; Mismatches 6; Indels 11; Gaps 2;  
 QY 2 YCG-----FLRPGVSS---ENLSAVATGNXG 24  
 DB 545 YCGWAPDGCIFLRPGTSATFEQDVSGASTSGLG 578  
 RESULT 4  
 SM6B RAT STANDARD; PRT; 887 AA.  
 AC 070141;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Semaphorin 6B precursor (Semaphorin 2) (Sema Z).  
 GN SEMA6B.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Wistar; TISSUE=Brain;  
 RX MEDLINE=98087397; PubMed=9427525;  
 RA Kikuchi K., Iehida H., Kimura T.;  
 RT "Molecular cloning of a novel member of semaphorin family genes,  
 semaphorin Z.";  
 RL Brain Res. Mol. Brain Res. 51:229-237(1997).

CC - FUNCTION: MAY PLAY A ROLE IN BOTH PERIPHERAL AND CENTRAL NERVOUS  
 CC SYSTEM DEVELOPMENT.  
 CC - SUBCELLULAR LOCATION: Type I membrane protein.  
 CC - SUBCELLULAR STAGE: DETECTED IN THE FIRST BRANCHIAL ARCH OF  
 CC EMBRYONIC DAY 11 (E11) EMBRYO, AND SUBSEQUENTLY IN THE MYOTOMES  
 CC AND THE DORSAL ROOT GANGLIA IN DEVELOPING SOMITES FROM E11.5  
 CC THROUGH E13.5, BUT NOT IN THE BRAIN. HOWEVER, AT E15, 18, 21 AND  
 CC P0, SEMAZ WAS HIGHLY EXPRESSED IN THE BRAIN.  
 CC - SIMILARITY: Belongs to the semaphorin family.  
 CC - SIMILARITY: Contains 1 Sema domain.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 EMBL; AB000776; BAA25687.1; -  
 InterPro; IPR003659; Plexin-like.  
 InterPro; IPR001627; Sema.  
 Pfam; PF01403; Sema; 1.  
 SMART; SM00423; PSI; 1.  
 SMART; SM00630; Sema; 1.  
 Signal; Transmembrane; Multigene family; Neurogenesis; Glycoprotein;  
 Developmental protein.  
 SIGNAL 1 26 POTENTIAL.  
 CHAIN 27 887 SEMAPHORIN 6B.  
 DOMAIN 27 605 EXTRACELLULAR (POTENTIAL).  
 TRANSMEM 606 626 POTENTIAL.  
 DOMAIN 627 887 CYTOPLASMIC (POTENTIAL).  
 DOMAIN 239 549 SEMA.  
 CARBOHYD 75 75 N-LINKED (GLCNAC. .) (POTENTIAL).  
 CARBOHYD 156 156 N-LINKED (GLCNAC. .) (POTENTIAL).  
 CARBOHYD 168 168 N-LINKED (GLCNAC. .) (POTENTIAL).  
 CARBOHYD 292 292 N-LINKED (GLCNAC. .) (POTENTIAL).  
 CARBOHYD 387 387 N-LINKED (GLCNAC. .) (POTENTIAL).  
 CARBOHYD 442 442 N-LINKED (GLCNAC. .) (POTENTIAL).  
 CARBOHYD 463 463 N-LINKED (GLCNAC. .) (POTENTIAL).  
 SEQUENCE 887 AA; 95752 MW; 09543F3F202CD301 CRC64;  
 Query Match 31.9%; Score 48.5; DB 1; Length 887;  
 Best Local Similarity 35.3%; Pred. No. 26;  
 Matches 12; Conservative 5; Mismatches 6; Indels 11; Gaps 2;  
 QY 2 YCG-----FLRPGVSS---ENLSAVATGNXG 24  
 DB 545 YCGWAPDGCIFLRPGTSATFEQDVSGASTSGLG 578  
 RESULT 5  
 SECY SYN3 STANDARD; PRT; 442 AA.  
 ID SECY SYN3  
 AC P77964;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Preprotein translocase secy subunit.  
 GN SECY OR SLL1814.  
 OS Synecocystis sp. (strain PCC 6803).  
 OC Bacteria; Cyanobacteria; Chroococcales; Synecocystis.  
 OX NCBI\_TaxID=11148;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=97061201; PubMed=8905231;  
 RA Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,  
 RA Miyajima N., Hirose M., Sugita M., Sasamoto S., Kimura T.,  
 RA Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Naruo K., Okumura S.,  
 RA Shimo S., Takeuchi C., Wada T., Watanabe A., Yamada M., Yasuda M.,  
 RA Tabata S.;  
 RT "Sequence analysis of the genome of the unicellular cyanobacterium  
 Synecocystis sp. strain PCC6803. II. Sequence determination of the

RT entire genome and assignment of potential protein-coding regions.";  
 RL DNA Res. 3:109-136(1996).  
 CC -|- FUNCTION: Involved in protein export. Interacts with secA and secE  
 CC to allow the translocation of proteins across the plasma membrane,  
 CC by forming part of a channel.  
 CC -|- SUBUNIT: One of seven secretory proteins (secA-F and secY) that  
 CC comprise the prokaryotic protein translocation apparatus.  
 CC -|- SUBCELLULAR LOCATION: Integral membrane protein.  
 CC -|- SIMILARITY: Belongs to the secY/SEC61-alpha family.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; D90905; BAAL7331.1; -;  
 CC PIR; S77484; S77484.  
 CC InterPro; IPR002208; SecY.  
 CC Pfam; PF00344; secY; 1.  
 CC PRINTS; PR00303; SECYTRNLCASE.  
 CC TIGRfam; TIGR00967; 3a0501s007; 1.  
 CC PROSITE; PS00755; SECY\_1; 1.  
 CC PROSITE; PS00756; SECY\_2; 1.  
 CC KW Protein transport; Translocation; Transmembrane; Complete proteome.  
 CC SEQUENCE 442 AA; 47999 MW; 641D89CF9C0655E5 CRC64;  
 CC  
 CC Query Match 31.6%; Score 48; DB 1; Length 442;  
 CC Best Local Similarity 56.2%; Pred. No. 15;  
 CC Matches 9; Conservative 2; Mismatches 5; Indels 0; Gaps 0;  
 CC  
 CC QY 14 NLSAVATGNXGCGAF 29  
 CC :|:|||||  
 CC Db 289 SLAGPATNGELGGFG 304  
 CC  
 CC RESULT 6  
 CC NU6M\_DUGDU  
 CC ID NU6M\_DUGDU STANDARD; PRT; 175 AA.  
 CC AC Q8W9M5;  
 CC DT 28-FEB-2003 (Rel. 41, Created)  
 CC DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 CC DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 CC DE NADH-ubiquinone oxidoreductase chain 6 (EC 1.6.5.3).  
 CC GN NTND6 OR ND6 OR NADH6.  
 CC OS Dugong dugon (Dugong).  
 CC OG Mitochondrion.  
 CC OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC OC Mammalia; Eutheria; Sirenia; Dugongidae; Dugong.  
 CC OX NCBI\_TaxID=29137;  
 CC RN [1]  
 CC RP SEQUENCE FROM N.A.  
 CC RX MEDLINE=22056029; PubMed=12034869;  
 CC RA Arnason U., Adegoke J.A., Rodin K., Born E.W., Esa Y.B., Gullberg A.,  
 CC RA Nilsson M., Short R.V., Xu X., Janke A.;  
 CC RT "Mammalian mitogenomic relationships and the root of the eutherian  
 CC tree.";  
 CC RL Proc. Natl. Acad. Sci. U.S.A. 99:8151-8156(2002).  
 CC -|- CATALYTIC ACTIVITY: NADH + ubiquinone = NAD(+) + ubiquinol.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; AJ421723; CAD18919.1; -;  
 CC InterPro; IPR001457; Oxidored\_q3.  
 CC Pfam; PF00499; oxidored\_q3; 1.

KW Oxidoreductase; NAD; Ubiquinone; Mitochondrion.  
 SQ SEQUENCE 175 AA; 18441 MW; 063F763F5F5E614A3 CRC64;  
 CC  
 CC Query Match 31.2%; Score 47.5; DB 1; Length 175;  
 CC Best Local Similarity 42.9%; Pred. No. 6.6;  
 CC Matches 15; Conservative 4; Mismatches 5; Indels 11; Gaps 3;  
 CC  
 CC QY 4 GELRPGVSSE-----NLSAVATGNXGCG---APG 29  
 CC :|:|||||  
 CC Db 15 GFV--GVSSKPSPVYGGGLVAGGVGGGIVWSFG 47  
 CC  
 CC RESULT 7  
 CC THD1\_LYCES  
 CC ID THD1\_LYCES STANDARD; PRT; 595 AA.  
 CC AC P25306;  
 CC DT 01-MAY-1992 (Rel. 22, Created)  
 CC DT 01-MAY-1992 (Rel. 22, Last sequence update)  
 CC DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 CC DE Threonine dehydratase biosynthetic, chloroplast precursor  
 CC DE (EC 4.3.1.19) (Threonine deaminase) (TD).  
 CC GN TD.  
 CC OS Lycopodium obscurum (Tomato).  
 CC OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 CC OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids;  
 CC OC Lamiales; Solanales; Solanaceae; Solanum.  
 CC OX NCBI\_TaxID=4081;  
 CC RN [1]  
 CC RP SEQUENCE FROM N.A., AND SEQUENCE OF 52-61.  
 CC RC STRAIN=cv. Tiny Tim LA154;  
 CC RX MEDLINE=91187855; PubMed=2011578;  
 CC RA Samach A., Hareven D., Gutfinger T., Ken-Dror S., Lifschitz E.;  
 CC RT "Biosynthetic threonine deaminase gene of tomato: isolation,  
 CC structure, and upregulation in floral organs.";  
 CC RL Proc. Natl. Acad. Sci. U.S.A. 88:2678-2682(1991).  
 CC -|- CATALYTIC ACTIVITY: L-threonine = 2-oxobutanate + NH(3).  
 CC -|- COFACTOR: Pyridoxal phosphate.  
 CC -|- ENZYME REGULATION: Allosterically inhibited by isoleucine.  
 CC -|- PATHWAY: Isoleucine biosynthesis; first step.  
 CC -|- SUBUNIT: Homotetramer.  
 CC -|- SUBCELLULAR LOCATION: Chloroplast.  
 CC -|- MISCELLANEOUS: Expression in mature flowers is increased over  
 CC 500-fold.  
 CC -|- SIMILARITY: Belongs to the serine/threonine dehydratase family.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; M61914; AAA34171.1; -;  
 CC DR EMBL; M61915; AAA68097.1; -;  
 CC DR PIR; A38628; A38628.  
 CC DR HSP; P04968; 1TDJ.  
 CC DR InterPro; IPR001926; B6 enzyme beta.  
 CC DR InterPro; IPR000634; S/T dehydratase BS.  
 CC DR InterPro; IPR005787; Thr\_dehydratase.  
 CC DR InterPro; IPR001721; ThrDh\_C.  
 CC DR Pfam; PF00291; PALP; 1.  
 CC DR Pfam; PF00585; Thr dehydrat C; 2.  
 CC DR TIGRfam; TIGR01124; ilvA 2Cterm; 1.  
 CC DR PROSITE; PS00165; DEHYDRATASE SER\_THR; 1.  
 CC KW Isoleucine biosynthesis; Lyase; Chloroplast; Pyridoxal phosphate;  
 CC FT TRANSIT 1 51 CHLOROPLAST.  
 CC FT CHAIN 52 595 THREONINE DEHYDRATASE BIOSYNTHETIC.  
 CC FT BINDING 143 143 PYRIDOXAL PHOSPHATE (BY SIMILARITY).  
 CC SEQUENCE 595 AA; 64937 MW; AC430B5DD9F0348 CRC64;  
 CC  
 CC Query Match 30.9%; Score 47; DB 1; Length 595;

modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

Best Local Similarity	42.9%;	Pred. No. 29;
Matches	9: Conservative	6: Mismatches
	4: Indels	2: Gaps

1 A YCGFLRPGVSSENLSAVATG 21

379 AYCEFFYK--IKNENIVAIASG 397

## RESULT 8

IMP3_HUMAN	STANDARD;	PTT;	307 AA.
ID	IMP3_HUMAN		
AC	Q96F15; Q96NE4; Q9NUK3;		
DT	15-MAR-2004 (Rel. 43, Created)		
DT	15-MAR-2004 (Rel. 43, Last sequence update)		
DT	15-MAR-2004 (Rel. 43, Last annotation update)		
DE	Immunity-associated nucleotide 4-like 1 protein (Immunity-associated		
DE	protein 3) (IAB-5).		
GN	IMAP3.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
OX	NCBI_Taxid=9606;		
RN	[1]		
RP	SEQUENCE FROM N A (ISOFORMS 1 AND 2).		

[I] SEQUENCE FROM N A (TSOFORMS 1 AND 2).

RC TISSUE=Lung, and Placenta;  
RA Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,  
RA Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H.,  
RA Takahashima M., Hosori T., Kakui Y., Kodaira H., Kondo Y., Sugawara M.,  
RA Watanabe S., Chiba Y., Ishida S., Murakawa K., Ono Y., Taguchi S.,  
RA Watanabe S., Kimura K., Murakami K., Iehii S., Kawai Y., Saito K.,  
RA Yamamoto J.J., Wakamatsu A., Nakamura Y., Nagahari K., Masuko Y.,  
RA Ninomiya K., Iwayanagi T.;  
RL "NEDO human cDNA sequencing project."  
RT Submitted (FEB-2000) to the EMBL/GenBank/DDBJ databases.

[2] SEQUENCE FROM N A (TSOFORM 1)

SEQUENCE-B-cell;  
MEDLINE=22398257; PubMed=12477932;  
RC Straubeberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.P., Bhat N.K.,  
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
RA Scapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Schetz T.E.,  
RA Brownstein M.J., Uesdin T.B., Toshiyuki S., Carninci P., Prange C.,  
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,  
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RA Richards S., Wozney K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
RA Villalon D.K., Murley D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Fahy J., Helton E., Kettman M., Madan A.C., Rodrigues S., Sanchez A.,  
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,  
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
RT "Generation and initial analysis of more than 15,000 full-length  
human and mouse cDNA sequences.";  
FT Proc Natl Acad Sci U S A 99:16899-16903(2002).  
RX

Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).

-I- SUBCELLULAR LOCATION: Type IV membrane protein. Mitochondrial

outer membrane (By similarity).

-|- ALTERNATIVE PRODUCTS:  
 = not alternative analysis. Named isoforms=2.

```
Event=Alternative splicing; Named isoforms=2;
```

```
Name=I;
TRACE -000E15-1. Science-Dien]aved:
```

```

NAME=2;
T80ID=Q96FI5-T; sequence=displayed;

```

```
Name=Z;  
  ISCID=08CE15-2; Sequence=VSP 008961;
```

TH01U=Q36F13-2; sequence=V01\_000002;  
Note-No experimental confirmation available;

-|- SIMILARITY: Belongs to the GTP-binding IAN family.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC

[illegible]

Query Match	Score 46.5;	DB 1;	Length 307;
Best Local Similarity	36.7%;	Pred. No. 17;	

**Figure 1**

Qy 4 GFLRPGVSSNLSA-----VATGNXGCG 26

## RESULT 9

MAAL_CLOTT	STANDARD;	PRT;	413 AA.
ID	MAAL_CLOTT		
AC	Q05514;		
DT	01-FEB-1995 (Rel. 31, Created)		
DT	01-FEB-1995 (Rel. 31, Last sequence update)		
DT	10-FEB-2003 (Rel. 42, Last annotation update)		
DT	10-OCT-2003 (Rel. 42, Last annotation update)		
DE	Methylaspartate ammonia-lyase (EC 4.3.1.2) (Beta-methylaspartase). ;		
OS	Clostridium tetanomorphum.		
OC	Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;		
OC	Clostridium.		
OX	NCBI_TaxID=1553;		
[1]	SEQUENCE OF 1-26.		
RP	SEQUENCE FROM N.A., AND SEQUENCE OF 1-26.		
RP	STRAIN=ATCC 15920 / DSM 528 / HI;		
RC	MEDLINE=93041773; PubMed=1420191;		
RC	MEDLINE=93041773; PubMed=8454064;		
RC	Goda S.K., Minton N.P., Botting N.P., Gani D.;		
RA	"Cloning, sequencing, and expression in <i>Escherichia coli</i> of the		
RA	Clostridium tetanomorphum gene encoding beta-methylaspartase and		
RT	characterization of the recombinant protein.";		
RT	Biochemistry 31:10747-10756(1992).		
RL	[2]		
RP	SEQUENCE OF 1-24 FROM N.A., AND SEQUENCE OF 1-24.		
RP	STRAIN=ATCC 15920 / DSM 528 / HI;		
RC	MEDLINE=93202282; PubMed=8454064;		
RC	MEDLINE=93202282; PubMed=8454064;		
RA	Brecht M., Kellermann J., Plueckthun A.;		
RA	"Cloning and sequencing of glutamate mutase component E from		
RT	Clostridium tetanomorphum.";		
RT	Clostridium tetanomorphum.";		
RT	FEBS Lett. 319:84-89(1993).		
RL	[3]		
RP	SEQUENCE OF 1-71 FROM N.A.		
RP	STRAIN=NCIMB 11547;		
RC	MEDLINE=93154518; PubMed=8428631;		
RC	MEDLINE=93154518; PubMed=8428631;		
RA	Holloway D.E., Marsh E.N.G.;		
RA	"Cloning and sequencing of glutamate mutase component E from		
RT	Clostridium tetanomorphum. Organization of the mut genes.";		
RT	FEBS Lett. 317:44-48(1993).		

```

[4]
RN  SEQUENCE OF 1-15.
RP  STRAIN=NCIMB 11547; PubMed=1397267;
RX  MEDLINE=93011908;
RA  Marsh E.N.G., Holloway D.E.;
RT  "Cloning and sequencing of glutamate mutase component S from
RT  Clostridium tetanomorphum. Homologies with other cobalamin-dependent
RT  enzymes";
RL  FES Lett. 310:167-170(1992).
CC  -!- CATALYTIC ACTIVITY: L-threo-3-methylaspartate = mesaconate +
CC  NH(3).
CC  -!- COFACTOR: Cobalamin.
CC  -!- PATHWAY: Fermentation of glutamate; second step.
CC  -!- SUBUNIT: Homodimer.
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
DR  EMBL; S48141; AAB24070.1; -.
DR  EMBL; X70499; CAA49911.1; -.
DR  EMBL; X70695; CAA50027.1; -.
DR  PIR; B44285; B44285.
DR  PDB; 1XCZ; 20-MAR-02.
DR  PDB; 1KD0; 20-MAR-02.
DR  InterPro; IPR006395; Met_Asp_am_lyase.
DR  Pfam; PF05034; MAAL; 1.
DR  TIGRFAMs; TIGR01502; B_methylAsp_ase; 1.
DR  Lysase; Vitamin B12: 3D-structure.
KW  SEQUENCE 413 AA; 45534 MW; 4451923DB035EF13 CRC64;
SQ
Query Match 30.6%; Score 46.5; DB 1; Length 413;
Best Local Similarity 50.0%; Pred. No. 23;
Matches 15; Conservative 1; Mismatches 7; Indels 7; Gaps 3;

QY 1 AYCGFLRPGVSSP-NLSAVATGNKG--CGA 27
DB 355 AYCG-----GTCNETRNSAEVTNIGMACGA 380

RESULT 10
Y877 MYCTU STANDARD; PRT; 262 AA.
ID Y877 MYCTU
AC Q10539;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Hypothetical protein Rv0877/MT0900/Mb0901.
GN Rv0877 OR MT0900 OR MTCY31.05 OR MB0901.
OS Mycobacterium tuberculosis, and
OS Mycobacterium bovis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773, 1765;
[1]
RN  SEQUENCE FROM N.A.
RP  SPECIES=M.tuberculosis; STRAIN=H37RV;
RX  MEDLINE=9825987; PubMed=9634230;
RA  Cole S.T., Brosch R., Parkhill J., Barry C.E. III, Tekala F.,
RA  Gordon S.V., Eiglmeier K., Gas S., Chillingworth T., Connor R.,
RA  Badcock K., Basham D., Brown D., Chillingworth T., Gentles S., Hamlin N., Holroyd S.,
RA  Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA  Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA  Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA  Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA  Sulston J.E., Taylor K., Whitehead S., Barrall B.G.;
RT  "Deciphering the biology of Mycobacterium tuberculosis from the
RT  complete genome sequence.";
RL  Nature 393:537-544(1998).

```

```

[2]
RN  SEQUENCE FROM N.A.
RP  SPECIES=M.tuberculosis; STRAIN=CDC 1551 / Oshkosh;
RX  MEDLINE=22206494; PubMed=12218036;
RA  Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA  Peterson J., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E.,
RA  Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,
RA  Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA  Bishai W., Jacobs W.R. Jr., Venter J.C., Fraser C.M.;
RT  "Whole-genome comparison of Mycobacterium tuberculosis clinical and
RT  laboratory strains.";
RL  J. Bacteriol. 184:5479-5490(2002).
CC  -----
RN  SEQUENCE FROM N.A.
RP  SPECIES=M.bovis; STRAIN=AF2122/97;
RX  MEDLINE=22709107; PubMed=12788972;
RA  Garnier T., Eiglmeier K., Camus J.-C., Medina N., Mansoor H.,
RA  Pryor M., Duthoy S., Grondin S., Lacroix C., Monsemp C., Simon S.,
RA  Harris B., Atkin R., Doggett J., Mayes R., Keating L., Wheeler P.R.,
RA  Parkhill J., Barrall B.G., Cole S.T., Gordon S.V., Hewinson R.G.;
RT  "The complete genome sequence of Mycobacterium bovis.";
RL  Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882(2003).
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
DR  EMBL; Z73101; CAA97384.1; -.
DR  EMBL; AE006977; AAK45142.1; -.
DR  EMBL; BX248336; CAD93762.1; -.
DR  PIR; B70780; B70780.
DR  TIGR; MT0900; -.
DR  Tuberculist; Rv0877; -.
KW  Hypothetical protein; Complete proteome.
SQ  SEQUENCE 262 AA; 27469 MW; 8466B819D3657DC3 CRC64;
Query Match 30.3%; Score 46; DB 1; Length 262;
Best Local Similarity 37.8%; Pred. No. 17;
Matches 14; Conservative 0; Mismatches 11; Indels 12; Gaps 2;

QY 3 CGFLRP-----GVSSENLSA---VATGNKXGCA 27
DB 194 CGFFLPLAGSLGAMFGVCGNELSADGHVVDVRYGCGA 230

RESULT 11
YGF0 YEAST STANDARD; PRT; 273 AA.
ID YGF0 YEAST
AC P53177;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Hypothetical 30.8 kDa protein in DUP2-TIF4632 intergenic region.
GN YGL050W.
OS Saccharomyces cerevisiae (Baker's Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
[1]
RN  SEQUENCE FROM N.A.
RP  STRAIN=S288C;
RX  MEDLINE=97377993; PubMed=9234674;
RA  Feuerhahn M., de Montigny J., Potier S., Souciet J.-L.;
RT  "The characterization of two new clusters of duplicated genes
RT  suggests a 'lego' organization of the yeast Saccharomyces cerevisiae
RT  chromosomes.";
RL  Yeast 13:861-869(1997).
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration

```

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----  
 CC EMBL; Z72572; CAA96752.1; -;  
 CC PIR; S64054; S64054.  
 CC GermOnline; 141098; -;  
 CC SGD; S0003018; YGL050W.  
 CC InterPro; IPR003827; DUF207.  
 CC Pfam; PF02676; DUF207; 1.  
 CC Hypothetical protein.  
 CC KW SEQUENCE 273 AA; 30806 MW; A36DG1F625B0D8EF CRC64;

Query Match 30.3%; Score 46; DB 1; Length 273;  
 Best Local Similarity 50.0%; Pred. No. 18;  
 Matches 8; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY 3 CGFLRPGVSSNLSAV 18  
 |||||:|:|:|:  
 DB 156 CGPREGSGNSNLVAI 171

## RESULT 12

CYA9 XENLA STANDARD; PRT; 1305 AA.  
 AC P98999;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Adenylyl cyclase, type IX (EC 4.6.1.1) (ATP pyrophosphate-lyase)  
 DE (Adenylyl cyclase) (xIAC).  
 GN ADCY9.  
 OS Xenopus laevis (African clawed frog).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae;  
 OC Xenopodinae; Xenopus.  
 OX NCBI\_TaxID=8355;  
 [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Ovary;  
 RX MEDLINE=97228584; PubMed=9074644;  
 RA Torrejon M., Echeverria V., Retamales G., Herrera L., Hinrichs M.V.,  
 RA Olate J.;  
 RT "Molecular cloning and expression of an adenylyl cyclase from Xenopus  
 RT laevis oocytes.";  
 RL FEBS Lett. 404:91-94 (1997).

CC -!- FUNCTION: May play a fundamental role in situations where fine  
 CC interplay between intracellular calcium and cAMP determines the  
 CC cellular function. May be a physiologically relevant docking site  
 CC for calcineurin (By similarity).  
 CC -!- CATALYTIC ACTIVITY: ATP = 3',5'-cyclic AMP + diphosphate.  
 CC -!- COFACTOR: Binds 2 magnesium ions per subunit.  
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein (By similarity).  
 CC -!- DOMAIN: Composed of two homologous domains.  
 CC -!- SIMILARITY: Belongs to the adenylyl cyclase class-4/guanylyl  
 CC cyclase family.  
 CC -!- SIMILARITY: Contains 1 FKBP-type pPIase domain.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; Z46958; CAA87082.1; -;  
 CC HSSP; P26769; IAB8.  
 CC InterPro; IPR001054; G\_cyclase.  
 CC Pfam; PF00211; guanylate\_cyc; 2.

DR SMART; SM00044; CYCC; 2.  
 DR PROSITE; PS00452; GUANYLATE\_CYCLASES\_1; 1.  
 DR PROSITE; PS0125; GUANYLATE\_CYCLASES\_2; 2.  
 KW lyase; cAMP biosynthesis; Transmembrane; Glycoprotein; Repeat;  
 KW Metal-binding; Magnesium; Isomerase; Rotamase.  
 FT DOMAIN 1 110  
 FT TRANSMEM 111 131  
 FT EXTRACELLULAR (POTENTIAL).  
 FT DOMAIN 132 134  
 FT TRANSMEM 135 155  
 FT POTENTIAL.  
 FT DOMAIN 156 164  
 FT CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 165 185  
 FT POTENTIAL.  
 FT DOMAIN 186 209  
 FT EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 210 229  
 FT POTENTIAL.  
 FT DOMAIN 230 235  
 FT CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 236 253  
 FT EXTRACELLULAR (POTENTIAL).  
 FT DOMAIN 254 274  
 FT POTENTIAL.  
 FT TRANSMEM 275 295  
 FT POTENTIAL.  
 FT DOMAIN 296 778  
 FT TRANSMEM 779 799  
 FT EXTRACELLULAR (POTENTIAL).  
 FT DOMAIN 800 810  
 FT POTENTIAL.  
 FT TRANSMEM 811 831  
 FT EXTRACELLULAR (POTENTIAL).  
 FT DOMAIN 832 859  
 FT POTENTIAL.  
 FT TRANSMEM 860 880  
 FT EXTRACELLULAR (POTENTIAL).  
 FT DOMAIN 881 904  
 FT POTENTIAL.  
 FT TRANSMEM 905 911  
 FT CYTOPLASMIC (POTENTIAL).  
 FT DOMAIN 912 932  
 FT EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 933 966  
 FT POTENTIAL.  
 FT TRANSMEM 967 987  
 FT CYTOPLASMIC (POTENTIAL).  
 FT DOMAIN 988 1305  
 FT CARBOHYD 199 199  
 FT CARBOHYD 946 946  
 FT CARBOHYD 951 951  
 FT METAL 393 393  
 FT METAL 394 394  
 FT SEQUENCE 1305 AA; 145410 MW; 8E1A5B79466B556A CRC64;

Query Match 30.3%; Score 46; DB 1; Length 1305;  
 Best Local Similarity 38.7%; Pred. No. 93;  
 Matches 12; Conservative 6; Mismatches 5; Indels 8; Gaps 2;

QY 6 LRPGVSSNLSAV-----ATGNKGC-----GAF 28  
 |||||:|:|:|:  
 DB 188 LNPGLGSDLSNLTFSTATSSCLSVQVGSF 218

## RESULT 13

YEEJ ECOLI STANDARD; PRT; 2358 AA.  
 ID YEEJ ECOLI  
 AC P76347; P94750;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Hypothetical protein yeeJ.  
 GN YEEJ OR B1978.  
 OS Escherichia coli.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Escherichia.  
 OX NCBI\_TaxID=562;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12 / MG1655;  
 RX MEDLINE=97426617; PubMed=9278503;  
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,  
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
 RA Mau B., Shao Y.;  
 RT "The complete genome sequence of Escherichia coli K-12.";  
 RL Science 277:1453-1474 (1997).





```

RC STRAIN=PPN-1;
RX MEDLINE=95204340; PubMed=7896697;
RA Schurr M.J., Vickrey J.F., Kumar A.P., Campbell A.L., Cunin R.,
RA Benjamin R.C., Shanley M.S., O'Donovan G.A.;
RT "Aspartate transcarbamoylase genes of Pseudomonas putida: requirement
RT for an inactive dihydroorotase for assembly into the dodecameric
RT holoenzyme.";
RL J. Bacteriol. 177:1751-1759(1995).
CC -|- FUNCTION: NON-FUNCTIONAL DHOASE.
CC -|- SUBUNIT: HETEROODICAMER OF 6 ACTIVE PYRB SUBUNITS AND 6 NON-
CC CATALYTIC PYRC' SUBUNITS.
CC -|- SIMILARITY: Belongs to the DHOase family. Pyc' subfamily.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M97254; AAA69779.1; -.
DR PIR; B56144; B56144.
DR HAMAP; MF_00220; atypical; 1.
DR InterPro; IPR006680; Amidohydro 1.
DR InterPro; IPR005847; Pept M38_reg.
DR Pfam; PF01979; Amidohydro 1; 1.
DR ProDom; PD000518; Urease; 1.
KW Pyrimidine biosynthesis.
SQ SEQUENCE 424 AA; 44264 MW; 87159A625340AC7C CRC64;

Query Match 29.9%; Score 45.5; DB 1; Length 424;
Best Local Similarity 44.4%; Pred. No. 34;
Matches 12; Conservative 3; Mismatches 11; Indels 1; Gaps 1;

Qy 4 GFLRPGVSENLS-AVATGNKCGAFG 29
Db 125 GALTKEGEQLAELVALRDTGCVAFG 151

```

Search completed: May 26, 2004, 18:41:19  
Job time : 2.46463 secs



GenCore version 5.1.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 26, 2004, 18:18:25 ; Search time 2.31439 Seconds  
(without alignments)  
3953.539 Million cell updates/sec

Title: US-09-302-812-12  
Perfect score: 152  
Sequence: 1 AYCGFLRPGVSSNLSAVATGNKGCAGFG 29

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SPTREMBL 25:\*

- 1: sp\_archaea:\*
- 2: sp\_bacteria:\*
- 3: sp\_fungi:\*
- 4: sp\_human:\*
- 5: sp\_invertebrate:\*
- 6: sp\_mammal:\*
- 7: sp\_mhc:\*
- 8: sp\_organelle:\*
- 9: sp\_phase:\*
- 10: sp\_plant:\*
- 11: sp\_rodent:\*
- 12: sp\_virus:\*
- 13: sp\_vertebrate:\*
- 14: sp\_unclassified:\*
- 15: sp\_rv:\*
- 16: sp\_bacterioph:\*
- 17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	149	98.0	976	4 Q9Y4W7	Q9Y4W7 homo sapien
2	149	98.0	976	4 Q86W56	Q86W56 homo sapien
3	149	98.0	976	4 Q7Z742	Q7Z742 homo sapien
4	149	98.0	977	6 O02776	O02776 bos taurus
5	144	94.7	920	11 Q8CB72	Q8CB72 mus musculus
6	144	94.7	961	11 Q80Y06	Q80Y06 mus musculus
7	144	94.7	968	11 Q86522	Q86522 mus musculus
8	139	91.4	972	11 Q9QM2	Q9QM2 rattus norv
9	69	45.4	723	5 Q96ON8	Q96ON8 drosophila
10	69	45.4	768	5 Q46043	Q46043 drosophila
11	60.5	39.8	543	2 Q33907	Q33907 shewanella
12	60	39.5	548	10 Q9SKB3	Q9SKB3 arabidopsis
13	57	37.5	522	10 Q8YVAL	Q8YVAL arabidopsis
14	55.5	36.5	513	16 Q8DF10	Q8DF10 vibrio vuln
15	55	36.2	578	5 Q9VIS9	Q9VIS9 drosophila
16	55	36.2	607	5 Q8MR68	Q8MR68 drosophila

17	54	35.5	358	16 Q8G798	Q8G798 bifidobacte
18	53	34.9	426	4 Q8NE20	Q8NE20 homo sapien
19	53	34.9	426	4 Q9HB89	Q9HB89 homo sapien
20	52	34.2	485	16 Q92Q80	Q92Q80 rhizobium m
21	52	34.2	494	11 Q9D2R3	Q9D2R3 mus musculus
22	52	34.2	495	11 Q8BZQ2	Q8BZQ2 mus musculus
23	51.5	33.9	849	12 Q9WQY5	Q9WQY5 bovine herp
24	51	33.6	379	16 Q8NQE8	Q8NQE8 corynebacte
25	51	33.6	644	16 Q98CV7	Q98CV7 rhizobium l
26	51	33.6	3716	16 Q50379	Q50379 mycobacteri
27	50.5	33.2	1384	16 Q8EUY9	Q8EUY9 mycoplasma
28	50	32.9	282	16 Q97G23	Q97G23 clostridium
29	50	32.9	618	13 Q8AVC4	Q8AVC4 xenopus lae
30	50	32.9	706	5 Q8SXG1	Q8SXG1 drosophila
31	50	32.9	759	11 Q9JUC9	Q9JUC9 rattus norv
32	50	32.9	861	5 Q9VRT8	Q9VRT8 drosophila
33	50	32.9	4019	5 Q9VJB4	Q9VJB4 drosophila
34	49.5	32.6	307	16 Q8ERZ3	Q8ERZ3 pseudomonas
35	49.5	32.6	504	16 Q9RYW7	Q9RYW7 deinococcus
36	49.5	32.6	764	5 Q19637	Q19637 caenorhabdi
37	49.5	32.6	781	5 Q867X0	Q867X0 caenorhabdi
38	49	32.2	175	10 Q9FI91	Q9FI91 arabidopsis
39	49	32.2	214	16 Q88IV4	Q88IV4 pseudomonas
40	49	32.2	362	4 Q8TC55	Q8TC55 homo sapien
41	49	32.2	385	11 Q8BMM6	Q8BMM6 mus musculus
42	49	32.2	495	5 Q9GQ43	Q9GQ43 giardia lam
43	49	32.2	548	5 Q9GQ45	Q9GQ45 giardia lam
44	49	32.2	568	16 Q9HXS5	Q9HXS5 pseudomonas
45	49	32.2	815	10 Q9SWB4	Q9SWB4 glycine max

ALIGNMENTS

RESULT 1

Q9Y4W7  
ID Q9Y4W7 PRELIMINARY; PRT; 976 AA.  
AC Q9Y4W7;  
DT 01-NOV-1999 (TrEMBLrel. 12, Created)  
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
DE Poly(ADP-ribose) glycohydrolase.  
GN HPARG.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=99380098; PubMed=10449915;  
RA Ame J.C., Apiau F., Jacobson E.L., Jacobson M.K.;  
RT "Assignment of the poly(ADP-ribose) glycohydrolase gene (PARG) to  
human chromosome 10q11.23 and mouse chromosome 14B by in situ  
hybridization"; Genet. 85:269-270 (1999).  
RL Cytogenet. Cell Genet. 85:269-270 (1999).  
DR EMBL; AF005043; AAB61614.1; -;  
DR Genbank; HGNC:8605; PARG.  
DR GO; GO:0005737; C:cytoplasm; TAS.  
DR GO; GO:0004649; F:poly(ADP-ribose) glycohydrolase activity; TAS.  
DR InterPro; IPR007724; PARG.  
DR Pfam; PF05028; PARG; 1.  
KW Hydrolase.  
SQ SEQUENCE 976 AA; 110903 MW; 42A0F51466FCE7DC CRC64;

Query Match 98.0%; Score 149; DB 4; Length 976;

Best Local Similarity 96.8%; Pred. No. 1e-12; Indels 0; Gaps 0;

Matches 28; Conservative 0; Mismatches 1;

QY 1 AYCGFLRPGVSSNLSAVATGNKGCAGFG 29  
|||||

Db 848 AYCGFLRPGVSSNLSAVATGNKGCAGFG 876  
|||||

```

RESULT 2
Q86W56 PRELIMINARY; PRT; 976 AA.
AC Q86W56;
DT 01-JUN-2003 (T-EMBLrel. 24, Created)
DT 01-JUN-2003 (T-EMBLrel. 24, Last sequence update)
DE 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
DE Similar to poly (ADP-ribose) glycohydrolase.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Skin;
RA Strausberg R.;
RL Submitted (APR-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC050560; A4H50560.1; -.
DR GO; GO:0016787; F:hydrolase activity; IEA.
DR InterPro; IPR007724; PARG.
DR Pfam; PF05028; PARG; 1.
KW Hydrolase.
SQ SEQUENCE 976 AA; 111109 MW; D6646353C6D0180E CRC64;

Query Match 98.0%; Score 149; DB 4; Length 976;
Best Local Similarity 96.6%; Pred. No. 1e-12;
Matches 28; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AYCGLRPGVSSSENLSAVATGNKCGAGF 29
DB 848 AYCGLRPGVSSSENLSAVATGNKCGAGF 876

RESULT 3
Q7Z742 PRELIMINARY; PRT; 976 AA.
AC Q7Z742;
DT 01-OCT-2003 (T-EMBLrel. 25, Created)
DT 01-OCT-2003 (T-EMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
DE Hypothetical protein.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Testis;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.D., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Sherman C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.P., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loguercio N.A., Peters G.J., Abramson R.D., Mullighy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalska U., Smailus D.E., Schnerk A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RA "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences."
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Testis;
RA Strausberg R.;

```

```

RL Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC052966; A4H52966.1; -.
KW Hypothetical protein.
SQ SEQUENCE 976 AA; 111090 MW; 47C8E81826C4A74A CRC64;

Query Match 98.0%; Score 149; DB 4; Length 976;
Best Local Similarity 96.6%; Pred. No. 1e-12;
Matches 28; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AYCGLRPGVSSSENLSAVATGNKCGAGF 29
DB 848 AYCGLRPGVSSSENLSAVATGNKCGAGF 876

RESULT 4
ID 002776 PRELIMINARY; PRT; 977 AA.
AC 002776;
DT 01-JUL-1997 (T-EMBLrel. 04, Created)
DT 01-JUL-1997 (T-EMBLrel. 04, Last sequence update)
DT 01-JUN-2003 (T-EMBLrel. 24, Last annotation update)
DE Poly (ADP-ribose) glycohydrolase.
GN BPARG.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97277328; PubMed=9115250;
RA Lin W., Ame J.C., Aboul-Ela N., Jacobson E.L., Jacobson M.K.;
RT "Isolation and characterization of the cDNA encoding bovine poly (ADP-ribose) glycohydrolase."
RL J. Biol. Chem. 272:11895-11901 (1997).
DR EMBL; U78975; AAB53370.1; -.
DR GO; GO:0016787; F:hydrolase activity; IEA.
DR InterPro; IPR007724; PARG.
DR Pfam; PF05028; PARG; 1.
KW Hydrolase.
SQ SEQUENCE 977 AA; 110837 MW; 87D2100F979DF377 CRC64;

Query Match 98.0%; Score 149; DB 6; Length 977;
Best Local Similarity 96.6%; Pred. No. 1e-12;
Matches 28; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AYCGLRPGVSSSENLSAVATGNKCGAGF 29
DB 849 AYCGLRPGVSSSENLSAVATGNKCGAGF 877

RESULT 5
Q8CB72 PRELIMINARY; PRT; 920 AA.
AC Q8CB72;
DT 01-MAR-2003 (T-EMBLrel. 23, Created)
DT 01-MAR-2003 (T-EMBLrel. 23, Last sequence update)
DT 01-JUN-2003 (T-EMBLrel. 24, Last annotation update)
DE Poly.
GN PARG.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Bone;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium,
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs."
RL Nature 420:563-573 (2002).

```

DR EMBL; AK036656; BAC29519.1; -.  
DR MGD; MGI:1347094; PARG.  
DR InterPro; IPR007724; PARG.  
DR Pfam; PF05028; PARG; 1.  
SQ SEQUENCE 920 AA; 103799 MW; 379E779E121B09A1 CRC64;  
Query Match 94.7%; Score 144; DB 11; Length 920;  
Best Local Similarity 93.1%; Pred. No. 5e-12; Indels 0; Gaps 0;  
Matches 27; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 1 AYCGFLRPGVSSNLSAVATGNKGGCGAFG 29  
DB 841 AYCGFLRPGVSSNLSAVATGNKGGCGAFG 869  
RESULT 6  
Q80YQ6 PRELIMINARY; PRT; 961 AA.  
AC Q80YQ6  
DT 01-JUN-2003 (TrEMBLrel. 24, Created)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Poly (ADP-ribose) glycohydrolase.  
OS Mus musculus (Mouse)  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6; TISSUE=Brain;  
RA Strausberg R.;  
RL Submitted (APR-2003) to the EMBL/GenBank/DBJ databases.  
DR EMBL; BC050892; AAH50892.1; -.  
DR GO; GO:0016787; F:hydrolase activity; IEA.  
DR InterPro; IPR007724; PARG.  
DR Pfam; PF05028; PARG; 1.  
KW Hydrolase.  
SQ SEQUENCE 961 AA; 108580 MW; 1558F120C367CF63 CRC64;  
Query Match 94.7%; Score 144; DB 11; Length 961;  
Best Local Similarity 93.1%; Pred. No. 5.3e-12;  
Matches 27; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 1 AYCGFLRPGVSSNLSAVATGNKGGCGAFG 29  
DB 841 AYCGFLRPGVSSNLSAVATGNKGGCGAFG 869  
RESULT 7  
O88622 PRELIMINARY; PRT; 968 AA.  
AC O88622  
DT 01-NOV-1998 (TrEMBLrel. 08, Created)  
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
DE Poly (ADP-ribose) glycohydrolase.  
GN PARG.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC MEDLINE=99380098; PubMed=10449915;  
RA Ame J.C., Aplou F., Jacobson E.L., Jacobson M.K.;  
RX "Assignment of the poly(ADP-ribose) glycohydrolase gene (PARG) to human chromosome 10q11.23 and mouse chromosome 14B by in situ hybridization."  
RT Cytogenet. Cell Genet. 85:269-270 (1999).  
RL EMBL; AF079557; AAC28735.1; -.  
DR MGD; MGI:1347094; PARG.  
DR GO; GO:0016787; F:hydrolase activity; IEA.  
DR InterPro; IPR007724; PARG.

DR Pfam; PF05028; PARG; 1.  
KW Hydrolase.  
SQ SEQUENCE 968 AA; 109169 MW; 398IDA4809E1E007 CRC64;  
Query Match 94.7%; Score 144; DB 11; Length 968;  
Best Local Similarity 93.1%; Pred. No. 5.3e-12;  
Matches 27; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 1 AYCGFLRPGVSSNLSAVATGNKGGCGAFG 29  
DB 840 AYCGFLRPGVSSNLSAVATGNKGGCGAFG 868  
RESULT 8  
Q9QYM2 PRELIMINARY; PRT; 972 AA.  
AC Q9QYM2  
DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
DE Poly (ADP-ribose) glycohydrolase.  
GN PARG.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
OX NCBI\_TaxID=10116;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BUF; TISSUE=Colon;  
RA Shimokawa T., Masutani M., Nagasawa S., Nozaki T., Ikota N., Araki S., Aoki Y., Nakagawa H., Sugimura T.;  
RT "Isolation and cloning of rat poly (ADP-ribose) glycohydrolase (Parg)."  
RT Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AB019366; BAA87901.1; -.  
DR GO; GO:0016787; F:hydrolase activity; IEA.  
DR InterPro; IPR007724; PARG.  
DR Pfam; PF05028; PARG; 1.  
KW Hydrolase.  
SQ SEQUENCE 972 AA; 109002 MW; 3DED7885684F59E9 CRC64;  
Query Match 91.4%; Score 139; DB 11; Length 972;  
Best Local Similarity 89.7%; Pred. No. 2.8e-11;  
Matches 26; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
QY 1 AYCGFLRPGVSSNLSAVATGNKGGCGAFG 29  
DB 844 AYCGFLRPGVSSNLSAVATGNKGGCGAFG 872  
RESULT 9  
Q96ON8 PRELIMINARY; PRT; 723 AA.  
AC Q96ON8  
DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
DE LD42380P.  
GN PARG OR EG:114E2.1 OR CG2864.  
OS Drosophila melanogaster (fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7227;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Berkely;  
RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J., Champagne M., Chavez C., Dorsett V., Farfan D., Frise E., George R., Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J., Nunoo J., Pacleb J., Paragas V., Park S., Phouanavong S., Wan K., Yu C., Lewis S.E., Rubin G.M., Celniker S.;  
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.

```

DR EMBL; AY051955; AAK93379.1; -.
DR FlyBase; FBgn0023216; PARG.
DR InterPro; IPR007724; PARG.
DR Pfam; PF05028; PARG; 1.
SQ SEQUENCE 723 AA; 81109 MW; 90A5BA53775D611C CRC64;

Query Match          45.4%; Score 69; DB 5; Length 723;
Best Local Similarity 48.6%; Pred. No. 0.27;
Matches 17; Conservative 1; Mismatches 3; Indels 14; Gaps 2;

QY 1 AYCGFLR-----PGVSSSENLSAVATGNKCGGAFG 29
   |||||
   |||||
Db 433 AYIGFVHMVTPPPG-----VATGNWCGGAFG 459

RESULT 10
O46043
ID O46043 PRELIMINARY; PRT; 768 AA.
AC O46043; 1998 (TrEMBLrel. 06, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE PARG protein.
GN PARG OR EG:114E2.1 OR CG2864.
OS Drosophila melanogaster (fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.

RC STRAIN=Berkley;
MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abrell J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballou R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.V., Benos P.V., Bertram B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K.J., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foeller C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Laoko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reiert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kianos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskaas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195 (2000).
RN [2]
RP SEQUENCE FROM N.A.

```

```

RA Ame J.-C., Jacobson M.K.;
RT "Isolation and characterization of the cDNA encoding Drosophila
RT poly(ADP-ribose) glycohydrolase.";
RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
[3]
RN SEQUENCE OF 46-768 FROM N.A.
RP Murphy L., Harris D., Barrell B.;
RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.
[4]
RN SEQUENCE OF 46-768 FROM N.A.
RP Benos P.;
RL Submitted (DEC-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AE003428; AAF45886.1; -.
DR EMBL; AF079556; AAC28734.1; -.
DR EMBL; Z98254; CAB10913.1; -.
DR FlyBase; FBgn0023216; PARG.
DR GO; GO:0016787; F.hydrolase activity; IEA.
DR InterPro; IPR007724; PARG.
DR Pfam; PF05028; PARG; 1.
KW Hydrolase.
SQ SEQUENCE 768 AA; 86288 MW; 2E620C42E643FC04 CRC64;

Query Match          45.4%; Score 69; DB 5; Length 768;
Best Local Similarity 48.6%; Pred. No. 0.29;
Matches 17; Conservative 1; Mismatches 3; Indels 14; Gaps 2;

QY 1 AYCGFLR-----PGVSSSENLSAVATGNKCGGAFG 29
   |||||
   |||||
Db 478 AYIGFVHMVTPPPG-----VATGNWCGGAFG 504

RESULT 11
O33907
ID O33907 PRELIMINARY; PRT; 543 AA.
AC O33907;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein.
OS Shewanella sp. SCRC-2738.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales;
OC Alteromonadaceae; Shewanella.
OX NCBI_TaxID=53560;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SCRC-2738;
RX MEDLINE=97419510; PubMed=9274025;
RA Takeyama H., Takeda D., Yazawa K., Yamada A., Matsumaga T.;
RT "Expression of the eicosapentaenoic acid synthesis gene cluster from
RT Shewanella sp. in a transgenic marine cyanobacterium, Synechococcus
RT sp.";
RL Microbiology 143:0-0(0).
DR EMBL; U73935; AAB81126.1; -.
DR PIR; T30186; T30186.
KW Hypothetical protein.
SQ SEQUENCE 543 AA; 59378 MW; 7233F53635B794C7 CRC64;

Query Match          39.8%; Score 60.5; DB 2; Length 543;
Best Local Similarity 46.7%; Pred. No. 3.3;
Matches 14; Conservative 4; Mismatches 9; Indels 3; Gaps 2;

QY 2 YCGFLRPGVSSSENLSAVATGNKCG--CGAFG 29
   |||:
   |||||
Db 99 YAGAMANGISSEEL-VIALGQILCGSFG 127

RESULT 12
O9SKB3
ID O9SKB3 PRELIMINARY; PRT; 548 AA.
AC O9SKB3; Q94E77;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)

```

```

DE Putative poly(ADP-ribose) glycohydrolase.
GN AT2G31870 OR TEJ.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; Core eudicots; rosids;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RA Lin X., Kaul S., Shea T.P., Fujii C.Y., Shen M., VanAken S.E.,
RA Barnstead M.E., Mason T.M., Bowman C.L., Renning C.M., Benito M.-I.,
RA Carrera A.J., Creasy T.H., Buell C.R., Town C.D., Niernan W.C.,
RA Fraser C.M., Venter J.C.;
RA Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
RL [2]
RP SEQUENCE FROM N.A.
RA Panda S., Poirier G.G., Kay S.A.;
RT "TEJ defines a role for poly-ADP-ribosylation in establishing period
RT length of the Arabidopsis circadian oscillator.";
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC006533; AAD3285.2; -
DR EMBL; AF394690; AAK72256.1; -
DR PIR; B84726; B84726
DR GO; GO:0016787; F:hydrolase activity; IEA.
DR InterPro; IPR007724; PARG.
DR Pfam; PF05028; PARG; 1.
DR Hydrolase.
KW SEQUENCE 548 AA; 62169 MW; FIA79FDA157C3329 CRC64;

Query Match 39.5%; Score 60; DB 10; Length 548;
Best Local Similarity 51.9%; Pred. No. 4;
Matches 14; Conservative 2; Mismatches 9; Indels 2; Gaps 1;

QY 5 FLRPGVSSSENL--SAVATGNKXGCGAFG 29
DB 432 FYVEGVNDEHDDGVATGNWCGVFG 458
| | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | |

RESULT 13
Q8VYAL PRELIMINARY; PRT; 522 AA.
AC Q8VYAL; 2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Putative poly(ADP-ribose) glycohydrolase.
GN AT2G31870.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; Core eudicots; rosids;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Nguyen M., Karlin-Neumann G., Southwick A., Lam B., Miranda M.,
RA Palm C.J., Bowser L., Jones T., Banh J., Carninci P., Chen H.,
RA Cheuk R., Chung M.K., Hayashizaki Y., Ishida J., Kamiya A., Kawai J.,
RA Kim C., Lin J., Liu S.X., Narusaka M., Pham P.K., Sakano H.,
RA Sakurai T., Satou M., Seki M., Shinn P., Yamada K., Shinozaki K.,
RA Ecker J., Theologis A., Davis R.W.;
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY072330; AAL61937.1; -
DR GO; GO:0016787; F:hydrolase activity; IEA.
DR InterPro; IPR007724; PARG.
DR Pfam; PF05028; PARG; 1.
DR Hydrolase.
KW SEQUENCE 522 AA; 58942 MW; 6FF98685D867CA15 CRC64;

Query Match 37.5%; Score 57; DB 10; Length 522;
Best Local Similarity 64.7%; Pred. No. 10;
Matches 11; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

```

```

QY 13 ENLSAVATGNKXGCGAFG 29
DB 424 EKKGIVATGNWCGVFG 440
| | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | |

RESULT 14
Q8DF10 PRELIMINARY; PRT; 513 AA.
AC Q8DF10; 2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Na+/alanine symporter.
GN VV10415.
OS Vibrio vulnificus.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
OC Vibrionaceae; Vibrio.
OX NCBI_TaxID=672;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CMCP6;
RA Rhee J.H., Kim S.Y., Chung S.S., Kim J.J., Moon Y.H., Jeong H.,
RA Choy H.E.;
RT "Complete genome sequence of Vibrio vulnificus CMCP6.";
RL Submitted (DEC-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AE016798; AAC08938.1; -
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0005283; P:sodium:amino acid transporter activity; IEA.
DR GO; GO:0006865; P:amino acid transport; IEA.
DR GO; GO:0006814; P:sodium ion transport; IEA.
DR InterPro; IPR002293; AA/rel_permease1.
DR InterPro; IPR001463; Na/Ala_sympt.
DR Pfam; PF01235; Na_Ala_sympt.
DR PRINTS; PRO0175; NAALASMPORT.
DR TIGRFAMs; TIGR00835; agcs; 1.
DR PROSITE; PS00873; NA_ALANINE_SYMP; 1.
DR Complete proteome.
KW SEQUENCE 513 AA; 55137 MW; 545B3948378CF10D CRC64;

Query Match 36.5%; Score 55.5; DB 16; Length 513;
Best Local Similarity 48.1%; Pred. No. 17;
Matches 13; Conservative 2; Mismatches 7; Indels 5; Gaps 1;

QY 3 CGFLRPGVSSSENL SAVATGNKXGCGAFG 29
DB 152 CGFLRPGVQNSI-----GNAVEAARG 173
| | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | |

RESULT 15
Q9VIS9 PRELIMINARY; PRT; 578 AA.
AC Q9VIS9; 2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE CG10366 protein.
GN CG10366.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkelley;
RA Adams M.D., Celnik S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazek R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Aghayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,

```

RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
RA Borkova D., Bouchan M.R., Bouck J., Brokstein P., Brottier P.,  
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,  
RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glaeser K.,  
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwan C.,  
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
RA Laoko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
RA Nelson D.R., Nelson K.A., Nixon K., Nusakern D.R., Pacleb J.M.,  
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,  
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
RT "The genome sequence of Drosophila melanogaster.";  
RL Science 287:2185-2195(2000).  
DR EMBL; AF003663; AAF53836.1; .;  
DR FlyBase; FBgn0032814; CGI0366.  
DR InterPro; IPR007087; Znf C2H2.  
DR Pfam; PF00096; zf-C2H2; 8.  
DR SMART; SM00355; Znf C2H2; 8.  
DR PROSITE; PS00028; ZINC\_FINGER\_C2H2\_1; 8.  
DR PROSITE; PS00157; ZINC\_FINGER\_C2H2\_2; 8.  
KW Metal-binding; Zinc; Zinc-finger.  
SQ SEQUENCE 578 AA; 65401 MW; 53F060B21E19010C CRC64;

Query Match 36.2%; Score 55; DB 5; Length 578;  
Best Local Similarity 57.9%; Pred. No. 22;  
Matches 11; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY 11 SSENLSAVATGNXGCGAFG 29  
: :||| ||| ||| |||  
Db 501 NKDNLSPVATKQSGCFAFG 519

Search completed: May 26, 2004, 18:46:10  
Job time : 4.31439 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 26, 2004, 17:19:25 ; Search time 1.26028 Seconds  
(without alignments)  
2466.140 Million cell updates/sec

Title: US-09-302-812-13  
Perfect score: 53  
Sequence: 1 FLINPELIVSR 11

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_29Jan04:\*  
1: Geneseq1980s:\*  
2: Geneseq1990s:\*  
3: Geneseq2000s:\*  
4: Geneseq2001s:\*  
5: Geneseq2002s:\*  
6: Geneseq2003as:\*  
7: Geneseq2003bs:\*  
8: Geneseq2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	53	100.0	11	5	Aae25636 Bovine po
2	53	100.0	11	5	Aau76027 Bovine po
3	53	100.0	11	5	Aau76018 Bovine po
4	53	100.0	11	6	Abg72285 Oligopept
5	53	100.0	968	5	Aae25631 Murine po
6	53	100.0	968	5	Aau76022 Mouse pol
7	53	100.0	968	5	Aau76013 Mouse pol
8	53	100.0	968	6	Abg72280 Mmurine p
9	53	100.0	977	5	Aae25629 Bovine po
10	53	100.0	977	5	Aau76020 Bovine po
11	53	100.0	977	5	Aau75799 Bovine po
12	53	100.0	977	6	Abg72278 Bovine po
13	52	98.1	976	5	Aae25630 Human pol
14	52	98.1	976	5	Aau76021 Human pol
15	52	98.1	976	5	Aau76012 Human pol
16	52	98.1	976	6	Abg72279 Human pol
17	47	88.7	10	5	Aae25650 Bovine po
18	41	77.4	249	6	Abu36785 Protein e
19	41	77.4	252	2	Aay35276 Chlamydia
20	37	69.8	103	3	Abb40810 Human ORF
21	37	69.8	103	5	Abp02512 Human ORF
22	37	69.8	1323	7	Aar5248 N-methyl-
23	37	69.8	1323	7	Add48372 Rat Prote
24	37	69.8	1324	7	Add48915 Rat Prote
25	37	69.8	1324	7	Add47929 Rat Prote

## ALIGNMENTS

### RESULT 1

AAE25636  
ID AAE25636 standard; peptide; 11 AA.

XX  
AC AAE25636;  
XX

DT 04-NOV-2002 (first entry)

DE Bovine poly adenosine diphosphate-ribose glycohydrolase oligopeptide #3.

XX  
KW Bovine; poly adenosine diphosphate-ribose glycohydrolase; bPARG; enzyme;  
KW ADP-ribose; neoplastic disorder; genotoxic oxidative stress; nontropic;  
KW cardiac disorder; neuronal disorder; reperfusion injury; neurotoxicity;  
KW Alzheimer's disease; Huntington's disease; Parkinson's disease; cardiant;  
KW cytostatic; vasotropic; neuroprotective; anticonvulsant; gene therapy;  
XX antisense therapy.  
XX Bos taurus.

XX  
FH Key Location/Qualifiers  
FT Misc-difference 23  
FT FT /label= Unknown  
XX

US6395543-B1.

28-MAY-2002.

23-FEB-2000; 2000US-00511507.

01-MAY-1998; 98US-0083768P.

30-APR-1999; 99US-00302812.

(KENT ) UNIV KENTUCKY RES FOUND.

Jacobson MK, Jacobson EL, Ame J, Lin W;

WPI; 2002-535641/57.

XX New nucleic acid molecule encoding bovine poly adenosine diphosphate-ribose glycohydrolase involved in cellular response to DNA damage, inhibition of which is useful for treating neoplastic disorders and neurodegenerative diseases.

XX Claim 10; Col 81-82; 77pp; English.

XX The invention relates to an isolated nucleic acid molecule which encodes a bovine poly (adenosine diphosphate (ADP)-ribose) glycohydrolase (bPARG) which catalyses release of ADP-ribose from an ADP-ribose polymer. The

CC invention is useful as probes and primer molecules that can be used in  
 CC hybridization assays and polymerase chain reaction (PCR) amplification.  
 CC The knowledge of the nucleotide sequence of the PARG gene permits the  
 CC preparation of antisense therapeutics containing sequences complementary  
 CC to the mRNA of PARG gene. The antisense therapeutics are useful to treat  
 CC neoplastic disorders and conditions caused by genotoxic oxidative stress  
 CC e.g., cardiac disorders, neuronal disorders, reperfusion injury,  
 CC neurotoxicity, Alzheimer's disease, Huntington's disease and Parkinson's  
 CC disease. The invention is useful in gene therapy and antisense therapy.  
 CC The present sequence is bPARG oligopeptide

XX Sequence 11 AA;  
 Query Match 100.0%; Score 53; DB 5; Length 11;  
 Best Local Similarity 100.0%; Pred. No. 0.00072;  
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FLINPELIVSR 11  
 |||||  
 Db 1 FLINPELIVSR 11

RESULT 2  
 AAU76027  
 ID AAU76027 standard; peptide; 11 AA.

XX AC AAU76027;  
 XX 08-MAY-2002 (first entry)  
 XX Bovine poly(ADP-ribose) glycohydrolase (PARG) oligopeptide 61.  
 XX  
 XX Bovine; poly(ADP-ribose) glycohydrolase; PARG; DNA repair; ADP-ribose;  
 XX adenosine diphosphate-ribose; neoplastic disorder; adenocarcinoma;  
 XX leukaemia; lymphoma; melanoma; myeloma; sarcoma; teratocarcinoma;  
 XX hyperplasia; hypertrophy; reperfusion; ischaemia; heart attack; stroke;  
 XX neurodegenerative disease; neurological disorder; Alzheimer's disease;  
 XX Huntington's disease; Parkinson's disease; oligopeptide 61.

XX OS Bos taurus.  
 XX USG337202-B1.  
 XX 08-JAN-2002.  
 XX 23-FEB-2000; 2000US-00511477.  
 XX 01-MAY-1998; 98US-0083768P.  
 XX 30-APR-1999; 99US-00302812.  
 XX (KENT ) UNIV KENTUCKY RES FOUND.  
 XX Jacobson MK, Jacobson EL, Ame J, Lin W;  
 XX WPI; 2002-163240/21.

XX Novel isolated and purified poly (ADP-ribose) glycohydrolase protein  
 XX which catalyses release of ADP-ribose from ADP ribose polymer, useful for  
 XX treating neoplastic and neurological disorders, heart attack and stroke.  
 XX Claim 4; Col 25; 81pp; English.

XX The present invention relates to a new poly(ADP-ribose) glycohydrolase  
 XX (PARG) protein which catalyses release of ADP-ribose from an ADP  
 XX (adenosine diphosphate)-ribose polymer. The PARG molecule of the  
 XX invention is useful for generating antibodies and can be inhibited or  
 XX activated for diagnosing and treating neoplastic disorders such as  
 XX adenocarcinoma, leukaemia, lymphoma, melanoma, myeloma, sarcoma,  
 XX teratocarcinoma, hyperplasia and hypertrophy, reperfusion following  
 XX ischaemia, heart attack, stroke, neurodegenerative diseases, neurological  
 XX disorders including Alzheimer's, Huntington's and Parkinson's diseases,  
 XX and related conditions. PARG levels may be enhanced to suppress DNA  
 XX repair and increase the cell's susceptibility to chemotherapy drugs.

CC Antagonists of PARG are administered to treat or prevent neoplastic  
 CC disorders. The present amino acid sequence represents bovine PARG  
 CC oligonucleotide 61. This peptide is one of several PARG oligopeptides  
 CC (AAU76025-AAU76028) of the invention

XX Sequence 11 AA;  
 Query Match 100.0%; Score 53; DB 5; Length 11;  
 Best Local Similarity 100.0%; Pred. No. 0.00072;  
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FLINPELIVSR 11  
 |||||  
 Db 1 FLINPELIVSR 11

RESULT 3  
 AAU76018  
 ID AAU76018 standard; peptide; 11 AA.

XX AC AAU76018;  
 XX 08-MAY-2002 (first entry)  
 XX Bovine poly(ADP-ribose) glycohydrolase (PARG) oligopeptide 61.  
 XX  
 XX Cow; poly(ADP-ribose) glycohydrolase; PARG; PARG modulator; ADP-ribose;  
 XX adenosine diphosphate-ribose; DNA repair; apoptosis; neoplasia; cancer;  
 XX inherited genetic disease; myocardial infarction; vascular stroke; aging;  
 XX neurodegeneration; Huntington's disease; Parkinson's disease;  
 XX Alzheimer's disease; neurotoxicity; oligopeptide 61.

XX OS Bos taurus.  
 XX USG333148-B1.  
 XX 25-DEC-2001.  
 XX 30-APR-1999; 99US-00302812.  
 XX 01-MAY-1998; 98US-0083768P.  
 XX (KENT ) UNIV KENTUCKY RES FOUND.  
 XX Jacobson MK, Jacobson EL, Ame J, Lin W;  
 XX WPI; 2002-153820/20.

XX Screening compounds for modulation of poly(ADP-ribose) glycohydrolase,  
 XX useful potentially for treating diseases associated with DNA damage, e.g.  
 XX cancer.  
 XX Claim 4; Col 25; 80pp; English.

XX The present invention relates to a new method for screening compounds for  
 XX ability to modulate activity of an enzyme that hydrolyses ADP (adenosine  
 XX diphosphate)-ribose from an ADP-ribose polymer. The compounds are  
 XX inhibitors or activators of PARG (poly(ADP-ribose) glycohydrolase) and  
 XX are used to treat or prevent any condition associated with DNA damage  
 XX e.g. neoplasia, inherited genetic diseases, myocardial infarction,  
 XX vascular stroke, aging and neurodegeneration e.g. Huntington's,  
 XX Parkinson's or Alzheimer's diseases, or neurotoxicity generally.  
 XX Compounds identified by the new method are more effective than known  
 XX inhibitors and have fewer side effects. The present amino acid sequence  
 XX represents bovine PARG oligopeptide 61. This peptide is one of several  
 XX PARG oligopeptides (AAU76016-AAU76019) of the invention

XX Sequence 11 AA;  
 Query Match 100.0%; Score 53; DB 5; Length 11;  
 Best Local Similarity 100.0%; Pred. No. 0.00072;  
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;



QY 1 FLINPELIVSR 11  
 DB 1 FLINPELIVSR 11

RESULT 4  
 ABG72285  
 ID ABG72285 standard; peptide; 11 AA.  
 AC ABG72285;  
 XX  
 XX 13-MAR-2003 (first entry)  
 XX  
 XX Oligopeptide #3 derived from bovine PARG enzyme.  
 XX  
 XX Bovine; poly(ADP-ribose) glycohydrolase; PARG; PARG expression;  
 KW cellular response; DNA damage; neoplastic disorder inducing agent;  
 KW oxidative stress; neoplastic disorder; myocardial infarction;  
 KW vascular stroke; neurodegenerative disorder; Alzheimer's disease;  
 KW Parkinson's disease; Huntington's disease; inborn genetic error;  
 KW reperfusion; ischaemia; aging; neurotoxicity; PARG activity; cytostatic;  
 KW neuroprotective; neurotropic; antiparkinsonian; cardiant; vasotrophic;  
 KW anticonvulsant; cerebroprotective.  
 XX  
 OS Bos taurus.  
 XX  
 XX US2002132328-A1.  
 XX  
 XX 19-SEP-2002.  
 XX  
 XX 09-OCT-2001; 2001US-00973451.  
 XX  
 XX 01-MAY-1998; 98US-0083768P.  
 XX 30-APR-1999; 99US-00302812.  
 XX  
 XX (JACO/) JACOBSON M K.  
 XX (JACO/) JACOBSON E L.  
 XX (AMEJ/) AME J.  
 XX (LINW/) LIN W.  
 XX  
 XX Jacobson MK, Jacobson EL, Ame J, Lin W;  
 XX WPI; 2003-155895/15.  
 XX  
 XX New nucleic acid molecule encoding a polypeptide with poly(ADP-ribose)  
 PT glycohydrolase activity, for preventing, treating, or ameliorating a  
 PT disease condition, e.g. neoplastic disorder, myocardial infarction or  
 PT vascular stroke.  
 XX  
 XX Example 2; Page 14; 86pp; English.  
 XX  
 XX The present invention relates to the isolation of poly(ADP-ribose)  
 CC glycohydrolase (PARG) from several species, and the polynucleotide  
 CC sequences encoding them. Methods for inhibiting PARG expression or  
 CC overexpressing PARG are also disclosed. PARG is involved in the cellular  
 CC response to DNA damage, and is associated with the body's response to  
 CC neoplastic disorder inducing agents and oxidative stress. The  
 CC polynucleotide sequences encoding PARG and PARG modulators are useful for  
 CC preventing, treating, or ameliorating diseases such as neoplastic  
 CC disorders, myocardial infarction, vascular stroke, neurodegenerative  
 CC disorders (e.g. Alzheimer's disease, Parkinson's disease or Huntington's  
 CC disease), inborn genetic errors, reperfusion following ischaemia, aging,  
 CC and neurotoxicity. The polynucleotide sequences are also useful in gene  
 CC therapy. The methods are useful for identifying an agent that modulates  
 CC PARG activity, identifying a mutant PARG allele in an individual, or  
 CC screening candidate molecules for PARG modulating activity. ABG72283-  
 CC ABG72286 represent oligopeptides derived from bovine PARG enzyme. The  
 CC oligopeptides are used to construct degenerate PCR primers for the  
 CC isolation of cDNA encoding bovine PARG

Best Local Similarity 100.0%; Pred. No. 0.00072;  
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FLINPELIVSR 11  
 DB 1 FLINPELIVSR 11

RESULT 5  
 AAE25631  
 ID AAE25631 standard; protein; 968 AA.  
 XX  
 XX AAE25631;  
 XX  
 XX 04-NOV-2002 (first entry)  
 XX  
 XX Murine poly adenosine diphosphate-ribose glycohydrolase (PARG).  
 XX  
 XX Murine; poly adenosine diphosphate-ribose glycohydrolase; PARG; enzyme;  
 KW ADP-ribose; neoplastic disorder; genotoxic oxidative stress; neurotropic;  
 KW cardiac disorder; neuronal disorder; reperfusion injury; neurotoxicity;  
 KW Alzheimer's disease; Huntington's disease; Parkinson's disease; cardiant;  
 KW cytostatic; vasotrophic; neuroprotective; anticonvulsant; gene therapy;  
 KW antisense therapy.  
 XX  
 XX Mus musculus.  
 XX  
 XX US6395543-B1.  
 XX  
 XX 28-MAY-2002.  
 XX  
 XX 23-FEB-2000; 2000US-00511507.  
 XX  
 XX 01-MAY-1998; 98US-0083768P.  
 XX 30-APR-1999; 99US-00302812.  
 XX  
 XX (KENT ) UNIV KENTUCKY RES FOUND.  
 XX  
 XX Jacobson MK, Jacobson EL, Ame J, Lin W;  
 XX WPI; 2002-535641/57.  
 XX N-PSDB; AAD42083.  
 XX  
 XX New nucleic acid molecule encoding bovine poly adenosine diphosphate-  
 PT ribose glycohydrolase involved in cellular response to DNA damage,  
 PT inhibition of which is useful for treating neoplastic disorders and  
 PT neurodegenerative diseases.  
 XX  
 XX Claim 3; Col 63-68; 77pp; English.  
 XX  
 XX The invention relates to an isolated nucleic acid molecule which encodes  
 CC a bovine poly (adenosine diphosphate (ADP)-ribose) glycohydrolase (bPARG)  
 CC which catalyses release of ADP-ribose from an ADP ribose polymer. The  
 CC invention is useful as probes and primer molecules that can be used in  
 CC hybridisation assays and polymerase chain reaction (PCR) amplification.  
 CC The knowledge of the nucleotide sequence of the PARG gene permits the  
 CC preparation of antisense therapeutics containing sequences complementary  
 CC to the mRNA of PARG gene. The antisense therapeutic are useful to treat  
 CC neoplastic disorders and conditions caused by genotoxic oxidative stress  
 CC e.g., cardiac disorders, neuronal disorders, reperfusion injury,  
 CC neurotoxicity, Alzheimer's disease, Huntington's disease and Parkinson's  
 CC disease. The invention is useful in gene therapy and antisense therapy.  
 XX  
 XX The present sequence is murine PARG

Query Match 100.0%; Score 53; DB 5; Length 968;  
 Best Local Similarity 100.0%; Pred. No. 0.14;  
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FLINPELIVSR 11  
 DB 751 FLINPELIVSR 761

XX	AAU76013;
XX	AC
XX	DT
XX	08-MAY-2002 (first entry)
XX	Mouse poly (ADP-ribose) glycohydrolase (PARG) protein sequence.
DE	Mouse; poly(ADP-ribose) glycohydrolase; PARG; PARG modulator; ADP-ribose;
XX	adenosine diphosphate-ribose; DNA repair; apoptosis; neoplasia; cancer;
KW	inherited genetic disease; myocardial infarction; vascular stroke; aging;
KW	neurodegeneration; Huntington's disease; Parkinson's disease;
KW	Alzheimer's disease; neurotoxicity.
OS	Mus musculus.
XX	OS
XX	US6333148-B1.
PX	PN
PD	25-DEC-2001.
PF	30-APR-1999; 99US-00302812.
PR	01-MAY-1998; 98US-0083768P.
XX	(KENT ) UNIV KENTUCKY RES FOUND.
XX	Jacobson MK, Jacobson BL, Ame J, Lin W;
XX	WPI; 2002-153820/20.
DR	N-PSDB; ABK14495.
XX	Screening compounds for modulation of poly(ADP-ribose) glycohydrolase,
PT	useful potentially for treating diseases associated with DNA damage, e.g.
PT	cancer.
XX	Claim 3; Col 63-68; 80pp; English.
XX	The present invention relates to a new method for screening compounds for
CC	ability to modulate activity of an enzyme that hydrolyses ADP (adenosine
CC	diphosphate)-ribose from an ADP-ribose polymer. The compounds are
CC	inhibitors or activators of PARG (poly(ADP-ribose) glycohydrolase) and
CC	are used to treat or prevent any condition associated with DNA damage
CC	e.g. neoplasia, inherited genetic diseases, myocardial infarction,
CC	vascular stroke, aging and neurodegeneration e.g. Huntington's,
CC	Parkinson's or Alzheimer's diseases, or neurotoxicity generally.
CC	Compounds identified by the new method are more effective than known
CC	inhibitors and have fewer side effects. The present amino acid sequence
CC	represents the mouse PARG protein of the invention. This protein is one
CC	of several PARG proteins (AAU75799 and AAU76012-AAU76015) of the
CC	invention
XX	SQ
XX	Sequence 968 AA;
Query Match	100.0%; Score 53; DB 5; Length 968;
Best Local Similarity	100.0%; Pred. No. 0.14;
Matches	11; Conservative 0; Mismatches 0; Indels 0; Gaps
Qy	1 FLINPELIVSR 11 
Db	751 FLINPELIVSR 761
RESULT 8	
ID	ABG72280
ID	ABG72280 standard; protein; 968 AA..
XX	AC
XX	ABG72280;
DT	13-MAR-2003 (first entry)
XX	Mmurine poly(ADP-ribose) glycohydrolase (PARG) enzyme.
DE	Murine; poly(ADP-ribose) glycohydrolase; PARG; PARG expression;
KW	cellular response; DNA damage; neoplastic disorder inducing agent;

KW oxidative stress; neoplastic disorder; myocardial infarction;  
 KW vascular stroke; neurodegenerative disorder; Alzheimer's disease;  
 KW Parkinson's disease; Huntington's disease; inborn genetic error;  
 KW reperfusion; ischemia; aging; neurotoxicity; PARG activity; mouse;  
 KW cytosolic; neuroprotective; neotropic; antiparkinsonian; cardiant;  
 KW vasotropic; anticonvulsant; cerebroprotective; enzyme.  
 XX  
 OS Mus musculus.  
 XX  
 XX US2002132328-A1.  
 XX  
 XX 19-SEP-2002.  
 XX  
 XX 09-OCT-2001; 2001US-00973451.  
 XX  
 XX 01-MAY-1998; 98US-0083768P.  
 XX  
 XX 30-APR-1999; 99US-00302812.  
 XX  
 XX (JACO/) JACOBSON M K.  
 XX (JACO/) JACOBSON E L.  
 XX (AMEJ/) AME J.  
 XX (LINW/) LIN W.  
 XX  
 XX Jacobson MK, Jacobson EL, Ame J, Lin W;  
 XX  
 XX WPI; 2003-155895/15.  
 XX N-PSDB; ABX14479.  
 XX  
 XX New nucleic acid molecule encoding a polypeptide with poly(ADP-ribose)  
 PT glycohydrolase activity, for preventing, treating, or ameliorating a  
 PT disease condition, e.g. neoplastic disorder, myocardial infarction or  
 PT vascular stroke.  
 XX  
 XX Claim 28; Fig 16; 86pp; English.  
 XX  
 XX The present invention relates to the isolation of poly(ADP-ribose)  
 CC glycohydrolase (PARG) from several species, and the polynucleotide  
 CC sequences encoding them. Methods for inhibiting PARG expression or  
 CC overexpressing PARG are also disclosed. PARG is involved in the cellular  
 CC response to DNA damage, and is associated with the body's response to  
 CC neoplastic disorder inducing agents and oxidative stress. The  
 CC polynucleotide sequences encoding PARG and PARG modulators are useful for  
 CC preventing, treating, or ameliorating diseases such as neoplastic  
 CC disorders, myocardial infarction, vascular stroke, neurodegenerative  
 CC disorders (e.g. Alzheimer's disease, Parkinson's disease or Huntington's  
 CC disease), inborn genetic errors, reperfusion following ischemia, aging,  
 CC and neurotoxicity. The polynucleotide sequences are also useful in gene  
 CC therapy. The methods are useful for identifying an agent that modulates  
 CC PARG activity, identifying a mutant PARG allele in an individual, or  
 CC screening candidate molecules for PARG modulating activity. The present  
 CC sequence represents murine PARG enzyme  
 XX  
 XX Sequence 968 AA;  
 Query Match 100.0%; Score 53; DB 6; Length 968;  
 Best Local Similarity 100.0%; Pred. No. 0.14;  
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 FLINPELIVSR 11  
 DB 751 FLINPELIVSR 761  
 |||||  
 RESULT 9  
 AA25629  
 ID AAE25629 standard; protein; 977 AA.  
 XX  
 XX AA25629;  
 XX  
 XX 04-NOV-2002 (first entry)  
 XX  
 XX Bovine poly adenosine diphosphate-ribose glycohydrolase (bPARG).  
 DE  
 XX

KW Bovine; poly adenosine diphosphate-ribose glycohydrolase; bPARG; enzyme;  
 KW ADP-ribose; neoplastic disorder; genotoxic oxidative stress; neotropic;  
 KW cardiac disorder; neuronal disorder; reperfusion injury; neurotoxicity;  
 KW Alzheimer's disease; Huntington's disease; Parkinson's disease; cardiant;  
 KW cytosolic; vasotropic; neuroprotective; anticonvulsant; gene therapy;  
 KW antisense therapy.  
 XX  
 OS Bos taurus.  
 XX  
 XX US6395543-B1.  
 XX  
 XX 28-MAY-2002.  
 XX  
 XX 23-FEB-2000; 2000US-00511507.  
 XX  
 XX 01-MAY-1998; 98US-0083768P.  
 XX  
 XX 30-APR-1999; 99US-00302812.  
 XX  
 XX (KENT) UNIV KENTUCKY RES FOUND.  
 XX  
 XX Jacobson MK, Jacobson EL, Ame J, Lin W;  
 XX  
 XX WPI; 2002-535641/57.  
 XX N-PSDB; AAD42081.  
 XX  
 XX New nucleic acid molecule encoding bovine poly adenosine diphosphate-  
 PT ribose glycohydrolase involved in cellular response to DNA damage,  
 PT inhibition of which is useful for treating neoplastic disorders and  
 PT neurodegenerative diseases.  
 XX  
 XX Claim 3; Col 47-45; 77pp; English.  
 XX  
 XX The invention relates to an isolated nucleic acid molecule which encodes  
 CC a bovine poly (adenosine diphosphate (ADP)-ribose) glycohydrolase (bPARG)  
 CC which catalyses release of ADP-ribose from an ADP-ribose polymer. The  
 CC invention is useful as probes and primer molecules that can be used in  
 CC hybridisation assays and polymerase chain reaction (PCR) amplification.  
 CC The knowledge of the nucleotide sequence of the PARG gene permits the  
 CC preparation of antisense therapeutics containing sequences complementary  
 CC to the mRNA of PARG gene. The antisense therapeutic are useful to treat  
 CC neoplastic disorders and conditions caused by genotoxic oxidative stress  
 CC e.g., cardiac disorders, neuronal disorders, reperfusion injury,  
 CC neurotoxicity, Alzheimer's disease, Huntington's disease and Parkinson's  
 CC disease. The invention is useful in gene therapy and antisense therapy.  
 CC The present sequence is bPARG  
 XX  
 XX Sequence 977 AA;  
 Query Match 100.0%; Score 53; DB 5; Length 977;  
 Best Local Similarity 100.0%; Pred. No. 0.15;  
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 FLINPELIVSR 11  
 DB 760 FLINPELIVSR 770  
 |||||  
 RESULT 10  
 AAU76020  
 ID AAU76020 standard; protein; 977 AA.  
 XX  
 XX AAU76020;  
 XX  
 XX 08-MAY-2002 (first entry)  
 XX  
 XX Bovine poly (ADP-ribose) glycohydrolase (PARG) protein sequence.  
 XX  
 XX Cow; poly (ADP-ribose) glycohydrolase; PARG; DNA repair; ADP-ribose;  
 KW adenosine diphosphate-ribose; neoplastic disorder; adenocarcinoma;  
 KW leukaemia; lymphoma; melanoma; myeloma; sarcoma; teratocarcinoma;  
 KW hyperplasia; hypertrophy; reperfusion; ischemia; heart attack; stroke;  
 KW neurodegenerative disease; neurological disorder; Alzheimer's disease;  
 KW Huntington's disease; Parkinson's disease.  
 KW



KW Bovine; poly(ADP-ribose) glycohydrolase; PARG; PARG expression;  
 KW cellular response; DNA damage; neoplastic disorder inducing agent;  
 KW oxidative stress; neoplastic disorder; myocardial infarction;  
 KW vascular stroke; neurodegenerative disorder; Alzheimer's disease;  
 KW Parkinson's disease; Huntington's disease; inborn genetic error;  
 KW reperfusion; ischaemia; aging; neurotoxicity; PARG activity; cytostatic;  
 KW neuroprotective; neotropic; antiparkinsonian; cardiant; vasotrophic;  
 KW anticonvulsant; cerebroprotective; enzyme.  
 XX  
 OS Bos taurus.  
 XX  
 XX US2002132328-A1.  
 XX  
 XX 19-SEP-2002.  
 XX  
 XX 09-OCT-2001; 2001US-00973451.  
 XX  
 XX 01-MAY-1998; 98US-0083768P.  
 XX 30-APR-1999; 99US-00302812.  
 XX  
 XX (JACO/) JACOBSON M K.  
 XX (JACO/) JACOBSON E L.  
 XX (AMEJ/) AME J.  
 XX (LINW/) LIN W.  
 XX  
 XX Jacobson MK, Jacobson EL, Ame J, Lin W;  
 XX  
 XX WPI; 2003-155895/15.  
 XX N-PSDB; ABX14477.  
 XX  
 XX New nucleic acid molecule encoding a polypeptide with poly(ADP-ribose)  
 PT glycohydrolase activity, for preventing, treating, or ameliorating a  
 PT disease condition, e.g. neoplastic disorder, myocardial infarction or  
 PT vascular stroke.  
 PT  
 PS Claim 28; Fig 16; 86pp; English.  
 PS  
 XX The present invention relates to the isolation of poly(ADP-ribose)  
 CC glycohydrolase (PARG) from several species, and the polynucleotide  
 CC sequences encoding them. Methods for inhibiting PARG expression or  
 CC overexpressing PARG are also disclosed. PARG is involved in the cellular  
 CC response to DNA damage, and is associated with the body's response to  
 CC neoplastic disorder inducing agents and oxidative stress. The  
 CC polynucleotide sequences encoding PARG and PARG modulators are useful for  
 CC preventing, treating, or ameliorating diseases such as neoplastic  
 CC disorders, myocardial infarction, vascular stroke, neurodegenerative  
 CC disorders (e.g. Alzheimer's disease, Parkinson's disease or Huntington's  
 CC disease), inborn genetic errors, reperfusion following ischaemia, aging,  
 CC and neurotoxicity. The polynucleotide sequences are also useful in gene  
 CC therapy. The methods are useful for identifying an agent that modulates  
 CC PARG activity, identifying a mutant PARG allele in an individual, or  
 CC screening candidate molecules for PARG modulating activity. The present  
 CC sequence represents bovine PARG enzyme  
 XX  
 XX Sequence 977 AA;  
 SQ  
 Query Match 100.0%; Score 53; DB 6; Length 977;  
 Best Local Similarity 100.0%; Pred. NO. 0.15; Indels 0; Gaps 0;  
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 FLINPELLIVSR 11  
 Db 760 FLINPELLIVSR 770  
 |||||  
 RESULT 13  
 AA25630  
 ID AA25630 standard; protein; 976 AA.  
 XX  
 AC AA25630;  
 XX  
 XX 04-NOV-2002 (first entry)  
 DT  
 XX

DE Human poly adenosine diphosphate-ribose glycohydrolase (PARG).  
 XX  
 XX Human; poly adenosine diphosphate-ribose glycohydrolase (PARG); enzyme;  
 KW ADP-ribose; neoplastic disorder; genotoxic oxidative stress; neotropic;  
 KW cardiac disorder; neuronal disorder; reperfusion injury; neurotoxicity;  
 KW Alzheimer's disease; Huntington's disease; Parkinson's disease; cardiant;  
 KW cytostatic; vasotrophic; neuroprotective; anticonvulsant; gene therapy;  
 KW antisenase therapy.  
 XX  
 XX Homo sapiens.  
 OS  
 XX US6395543-B1.  
 XX  
 XX 28-MAY-2002.  
 XX  
 XX 23-FEB-2000; 2000US-00511507.  
 XX  
 XX 01-MAY-1998; 98US-0083768P.  
 XX 30-APR-1999; 99US-00302812.  
 XX  
 XX (KENT ) UNIV KENTUCKY RES FOUND.  
 XX  
 XX Jacobson MK, Jacobson EL, Ame J, Lin W;  
 XX  
 XX WPI; 2002-535641/57.  
 XX N-PSDB; AAD42082.  
 XX  
 XX New nucleic acid molecule encoding bovine poly adenosine diphosphate-  
 PT ribose glycohydrolase involved in cellular response to DNA damage,  
 PT inhibition of which is useful for treating neoplastic disorders and  
 PT neurodegenerative diseases.  
 PT  
 XX Claim 3; Col 55-60; 77pp; English.  
 XX  
 XX The invention relates to an isolated nucleic acid molecule which encodes  
 CC a bovine poly (adenosine diphosphate (ADP)-ribose) glycohydrolase (PARG)  
 CC which catalyses release of ADP-ribose from an ADP-ribose polymer. The  
 CC invention is useful as probes and primer molecules that can be used in  
 CC hybridization assays and polymerase chain reaction (PCR) amplification.  
 CC The knowledge of the nucleotide sequence of the PARG gene permits the  
 CC preparation of antisense therapeutics containing sequences complementary  
 CC to the mRNA of PARG gene. The antisense therapeutics are useful to treat  
 CC neoplastic disorders and conditions caused by genotoxic oxidative stress  
 CC e.g., cardiac disorders, neuronal disorders, reperfusion injury,  
 CC neurotoxicity, Alzheimer's disease, Huntington's disease and Parkinson's  
 CC disease. The invention is useful in gene therapy and antisense therapy.  
 CC  
 XX The present sequence is human PARG  
 XX  
 XX Sequence 976 AA;  
 SQ  
 Query Match 98.1%; Score 52; DB 5; Length 976;  
 Best Local Similarity 90.9%; Pred. NO. 0.23;  
 Matches 10; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 FLINPELLIVSR 11  
 Db 759 FLINPELLIVSR 769  
 |||||  
 RESULT 14  
 AAU76021  
 ID AAU76021 standard; protein; 976 AA.  
 XX  
 XX AAU76021;  
 XX  
 XX 08-MAY-2002 (first entry)  
 DT  
 XX Human poly(ADP-ribose) glycohydrolase (PARG) protein sequence.  
 DE  
 XX Human; poly(ADP-ribose) glycohydrolase; PARG; DNA repair; ADP-ribose;  
 KW adenosine diphosphate-ribose; neoplastic disorder; adenocarcinoma;  
 KW leukaemia; lymphoma; melanoma; myeloma; sarcoma; teratocarcinoma;  
 KW hyperplasia; hypertrophy; reperfusion; ischaemia; heart attack; stroke;

Thu May 27 09:55:57 2004

us-09-302-812-13.rag

KW neurodegenerative disease; neurological disorder; Alzheimer's disease;  
KW Huntington's disease; Parkinson's disease.

OS Homo sapiens.

XX US6337202-B1.

XX 08-JAN-2002.

XX 23-FEB-2000; 2000US-00511477.

XX 01-MAY-1998; 98US-0083768P.

XX 30-APR-1999; 99US-00302812.

XX (KENT ) UNIV KENTUCKY RES FOUND.

XX Jacobson MK, Jacobson EL, Ame J, Lin W;

XX WPI; 2002-163240/21.

XX N-PSDB; ABK14932.

XX Novel isolated and purified poly (ADP-ribose) glycohydrolase protein  
XX which catalyses release of ADP-ribose from ADP ribose polymer, useful for  
XX treating neoplastic and neurological disorders, heart attack and stroke.

XX Claim 2; Col 55-60; 81pp; English.

XX The present invention relates to a new poly(ADP-ribose) glycohydrolase  
XX (PARG) protein which catalyses release of ADP-ribose from an ADP  
XX (adenosine diphosphate)-ribose polymer. The PARG molecule of the  
XX invention is useful for generating antibodies and can be inhibited or  
XX activated for diagnosing and treating neoplastic disorders such as  
XX adenocarcinoma, leukaemia, lymphoma, melanoma, myeloma, sarcoma,  
XX teratocarcinoma, hyperplasia and hypertrophy, reperfusion following  
XX ischaemia, heart attack, stroke, neurodegenerative diseases, neurological  
XX disorders including Alzheimer's, Huntington's and Parkinson's diseases,  
XX and related conditions. PARG levels may be enhanced to suppress DNA  
XX repair and increase the cell's susceptibility to chemotherapy drugs.  
XX Antagonists of PARG are administered to treat or prevent neoplastic  
XX disorders. The present amino acid sequence represents the human PARG  
XX protein of the invention. This protein is one of several PARG proteins  
XX (AAU76020-AAU76024) of the invention

XX Sequence 976 AA;

Query Match 98.1%; Score 52; DB 5; Length 976;

Best Local Similarity 90.9%; Pred. No. 0.23;

Matches 10; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 FLINPELIVSR 11

|||||:|

Db 759 FLINPELLIISR 769

RESULT 15

AAU76012

ID AAU76012 standard; protein; 976 AA.

XX AC AAU76012;

XX 08-MAY-2002 (first entry)

XX Human poly(ADP-ribose) glycohydrolase (PARG) protein sequence.

XX Human; poly(ADP-ribose) glycohydrolase; PARG; PARG modulator; ADP-ribose;  
XX adenosine diphosphate-ribose; DNA repair; apoptosis; neoplasia; cancer;  
XX inherited genetic disease; myocardial infarction; vascular stroke; aging;  
XX neurodegeneration; Huntington's disease; Parkinson's disease;  
XX Alzheimer's disease; neurotoxicity.

XX Homo sapiens.

XX US6333148-B1.

XX 25-DEC-2001.

XX 30-APR-1999; 99US-00302812.

XX 01-MAY-1998; 98US-0083768P.

XX (KENT ) UNIV KENTUCKY RES FOUND.

XX Jacobson MK, Jacobson EL, Ame J, Lin W;

XX WPI; 2002-153820/20.

XX N-PSDB; ABK14494.

XX Screening compounds for modulation of poly(ADP-ribose) glycohydrolase,  
XX useful potentially for treating diseases associated with DNA damage, e.g.  
XX cancer.

XX Claim 3; Col 55-60; 80pp; English.

XX The present invention relates to a new method for screening compounds for  
XX ability to modulate activity of an enzyme that hydrolyses ADP (adenosine  
XX diphosphate)-ribose from an ADP-ribose polymer. The compounds are  
XX inhibitors or activators of PARG (poly(ADP-ribose) glycohydrolase) and  
XX are used to treat or prevent any condition associated with DNA damage,  
XX e.g. neoplasia, inherited genetic diseases, myocardial infarction,  
XX vascular stroke, aging and neurodegeneration e.g. Huntington's,  
XX Parkinson's or Alzheimer's diseases, or neurotoxicity generally.  
XX Compounds identified by the new method are more effective than known  
XX inhibitors and have fewer side effects. The present amino acid sequence  
XX represents the human PARG protein of the invention. This protein is one  
XX of several PARG proteins (AAU75799 and AAU76012-AAU76015) of the  
XX invention

XX Sequence 976 AA;

Query Match 98.1%; Score 52; DB 5; Length 976;

Best Local Similarity 90.9%; Pred. No. 0.23;

Matches 10; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 FLINPELIVSR 11

|||||:|

Db 759 FLINPELLIISR 769

Search completed: May 26, 2004, 18:40:12

Job time : 4.26028 secs

Result No.	Score		Query Match	Length	DB	ID	Description
	Score	Match					
1	516.4	100.0	968	11	Q88622	mus musculus	
2	5098.5	98.7	961	11	Q80Y06	mus musculus	
3	4869.5	94.3	920	11	Q8CB72	mus musculus	
4	4749	92.0	972	11	Q3QYW2	rattus norv	
5	4456	86.3	976	4	Q8W56	homo sapien	
6	4451	86.2	976	4	Q72742	homo sapien	
7	4395	85.1	976	4	Q9Y4W7	homo sapien	
8	4332.5	83.9	977	6	O02776	bos taurus	
9	1074.5	20.8	768	5	Q46043	drosochila	
10	1069.5	20.7	723	5	Q96ON8	drosochila	
11	568	11.0	548	10	Q9SKB3	arabidopsis	
12	525	10.2	522	10	Q8VY11	arabidopsis	
13	458	8.9	485	5	Q9NSL4	caenorhabdi	
14	451	8.7	781	5	Q867X0	caenorhabdi	
15	450.5	8.7	764	5	Q19637	caenorhabdi	
16	328.5	6.4	368	5	Q86G14	toxoplasma	

Db	61	RAGPHRGNAISFVFKQKTIITWMDTKGPKTAESKKNNTTRIDSMSSVQKDNFYPHK	120
Qy	121	EKLENVQNLNDKSTPESKSOYLNOQOQASVCKWQNEKGHAEQOLLAEPAGTLPKQLS	180
Db	121	EKLENVQNLNDKSTPESKSOYLNOQOQASVCKWQNEKGHAEQOLLAEPAGTLPKQLS	180
Qy	181	NANIGQSPHTDHSDDTDEEDRDNQOFTPIKLANTKPTVGDQARSCKSGSRQSKD	240
Db	181	NANIGQSPHTDHSDDTDEEDRDNQOFTPIKLANTKPTVGDQARSCKSGSRQSKD	240
Qy	241	CTGCOOEVDVLPESPLSDVGAEDIGTGPKNNDKLTGQESSLGDSPPEKSESPMDV	300
Db	241	CTGCOOEVDVLPESPLSDVGAEDIGTGPKNNDKLTGQESSLGDSPPEKSESPMDV	300
Qy	301	DNSNRSCDSEADETSPVDFDQDRSSQTANKLSSCOAREADGDLRKRYLTGSEVRH	360
Db	301	DNSNRSCDSEADETSPVDFDQDRSSQTANKLSSCOAREADGDLRKRYLTGSEVRH	360
Qy	361	FOFEGENAGTSDLNAPSGNSSSLNVECRSKQHGKDSKITDHFMRISKEDRRKQC	420
Db	361	FOFEGENAGTSDLNAPSGNSSSLNVECRSKQHGKDSKITDHFMRISKEDRRKQC	420
Qy	421	EVRHORTEKIPIKYPNLPPEKKMLGTPIEBMRKMPRCGIHLPSLRPSASHTVTRVDL	480
Db	421	EVRHORTEKIPIKYPNLPPEKKMLGTPIEBMRKMPRCGIHLPSLRPSASHTVTRVDL	480
Qy	481	LRAGEVPPFPFTHYKDLNDKHWKMPCSQNLNYPVEDENGERTAGSRWELIOTALLNKF	540
Db	481	LRAGEVPPFPFTHYKDLNDKHWKMPCSQNLNYPVEDENGERTAGSRWELIOTALLNKF	540
Qy	541	RPQNLKDALIKYNVASKWDFALTALVDFWVKLEAEQAHLQOSILPDMVKIALCLPNC	600
Db	541	RPQNLKDALIKYNVASKWDFALTALVDFWVKLEAEQAHLQOSILPDMVKIALCLPNC	600
Qy	601	TOPIPLKQKMHVSVTMSQEQIASILLANAFCTFPRRNAMKSEYSSYPDINFNRLFEG	660
Db	601	TOPIPLKQKMHVSVTMSQEQIASILLANAFCTFPRRNAMKSEYSSYPDINFNRLFEG	660
Qy	661	SSRKEPKLTLFCYFRVTEKPTGLVTFTRQSLDPEPWERCEKPLTLRHVYEGTIEG	720
Db	661	SSRKEPKLTLFCYFRVTEKPTGLVTFTRQSLDPEPWERCEKPLTLRHVYEGTIEG	720
Qy	721	NGRGLMLQVDFANRFVGGVGTAGLVOEIRFLINPELIVSRFLTEVLDHNECLITGTQE	780
Db	721	NGRGLMLQVDFANRFVGGVGTAGLVOEIRFLINPELIVSRFLTEVLDHNECLITGTQE	780
Qy	781	YSEYTGATYRWARSHEDGSEKDDWQRCTEIVADLHFRYLDQFVPEKVRRELNA	840
Db	781	YSEYTGATYRWARSHEDGSEKDDWQRCTEIVADLHFRYLDQFVPEKVRRELNA	840
Qy	841	YCGFLRPGVPSNLAVATGNMCGGAFGGDARUKALIQILAAAAAARDVYFTFGSELM	900
Db	841	YCGFLRPGVPSNLAVATGNMCGGAFGGDARUKALIQILAAAAAARDVYFTFGSELM	900
Qy	901	RDYSHMTFLTERKLDGVKVKLLRYNNEECNCSTPGDILKYPFIHAYESSAETTD	960
Db	901	RDYSHMTFLTERKLDGVKVKLLRYNNEECNCSTPGDILKYPFIHAYESSAETTD	960
Qy	961	MPGQKAGT 968	
Db	961	MPGQKAGT 968	

RESULT 2

Q80YQ6

ID Q80YQ6 PRELIMINARY; PRT; 961 AA.

AC Q80YQ6;

DT 01-JUN-2003 (TEMBLrel. 24, Created)

DT 01-JUN-2003 (TEMBLrel. 24, Last sequence update)

DT 01-OCT-2003 (TEMBLrel. 25, Last annotation update)

DE Poly (ADP-ribose) glycohydrolase.

OS Mus musculus (Mouse), Chordata; Craniata; Vertebrata; Euteleostomi;

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
OX	NCBI_TaxID=10090;		
RN	[L]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=C57BL/6; TISSUE=Brain;		
RA	Strausberg R.;		
RL	Submitted (APR-2003) to the EMBL/GenBank/DBJ databases.		
DR	EMBL; BC050892; AAH50892.1; -		
DR	GO; GO:0016787; F:hydrolase activity; IEA.		
DR	InterPro; IPR007724; PARG.		
DR	Pfam; PF05028; PARG; 1.		
KW	Hydrolase.		
SQ	SEQUENCE 961 AA; 108580 MW; 1558F120C367CF63 CRC64;		
Query Match 98.7%; Score 5098.5; DB 11; Length 961;			
Best Local Similarity 99.6%; Pred. No. 0;			
Matches 957; Conservative 1; Mismatches 2; Indels 1; Gaps 1;			
Qy	1	MSAGPWEPCCTK-ARWGAAGTSAPTASDSRSFQQRVRLDPKQAPVQFRVPPSSPACVS	59
Db	1	MSAGPWEPCCTKPRWGAAGTSAPTASDSRSFQQRVRLDPKQAPVQFRVPPSSPACVS	60
Qy	60	GRAGPHRGNAISFVFKQKTIITWMDTKGPKTAESKKNNTTRIDSMSSVQKDNFYPHK	119
Db	61	GRAGPHRGNAISFVFKQKTIITWMDTKGPKTAESKKNNTTRIDSMSSVQKDNFYPHK	120
Qy	120	VEKLENVQNLNDKSTPESKSOYLNOQOQASVCKWQNEKGHAEQOLLAEPAGTLPKQL	179
Db	121	VEKLENVQNLNDKSTPESKSOYLNOQOQASVCKWQNEKGHAEQOLLAEPAGTLPKQL	180
Qy	180	SNANIGQSPHTDHSDDTDEEDRDNQOFTPIKLANTKPTVGDQARSCKSGSRQSVK	239
Db	181	SNANIGQSPHTDHSDDTDEEDRDNQOFTPIKLANTKPTVGDQARSCKSGSRQSVK	240
Qy	240	DCTGCOOEVDVLPESPLSDVGAEDIGTGPKNDNKLTGQESSLGDSPPEKSESPSPMD	299
Db	241	DCTGCOOEVDVLPESPLSDVGAEDIGTGPKNDNKLTGQESSLGDSPPEKSESPSPMD	300
Qy	300	VDSNRNSCODSEADETSPVDFDQDRSSQTANKLSSCOAREADGDLRKRYLTGSEVRL	359
Db	301	VDSNKSNCODSEADETSPVDFDQDRSSQTANKLSSCOAREADGDLRKRYLTGSEVRL	360
Qy	360	HFQFEGENAGTSDLNAPSGNSSSLNVECRSSKHGKDRDSKITDHFWRISKSDRRKEQ	419
Db	361	HFQFEGENAGTSDLNAPSGNSSSLNVECRSSKHGKDRDSKITDHFWRISKSDRRKEQ	420
Qy	420	CEVRHORTEKIPKIYPNLPPEKKMLGTPIEEMRKMPRCGIHLPSLRPSASHTVTVRD	479
Db	421	CEVRHORTEKIPKIYPNLPPEKKMLGTPIEEMRKMPRCGIHLPSLRPSASHTVTVRD	480
Qy	480	LLRAGEVPKPPFTHYKDLNDKHWKMPCSEONLVPVEDENGERTAGSRWELIQTALLNKF	539
Db	481	LLRAGEVPKPPFTHYKDLNDKHWKMPCSEONLVPVEDENGERTAGSRWELIQTALLNKF	540
Qy	540	TRPQNLKDALIKYNVAYSKKWDFALTALVDFWMDVKLEAEQAHLQOSILPDMVKIALCLPNI	599
Db	541	TRPQNLKDALIKYNVAYSKKWDFALTALVDFWMDVKLEAEQAHLQOSILPDMVKIALCLPNI	600
Qy	600	CTQPIPLKQKMHVSVTMSQEQIASILLANAFCTFPRRNAMKSEYSSYPDINFNRLFEG	659
Db	601	CTQPIPLKQKMHVSVTMSQEQIASILLANAFCTFPRRNAMKSEYSSYPDINFNRLFEG	660
Qy	660	RSSRKEPKLTLFCYFRVTEKPTGLVTFTRQSLDPEPWERCEKPLTLRHVYEGTIE	719
Db	661	RSSRKEPKLTLFCYFRVTEKPTGLVTFTRQSLDPEPWERCEKPLTLRHVYEGTIE	720
Qy	720	NGRGLMLQVDFANRFVGGVGTAGLVOEIRFLINPELIVSRFLTEVLDHNECLITGTQE	779
Db	721	NGRGLMLQVDFANRFVGGVGTAGLVOEIRFLINPELIVSRFLTEVLDHNECLITGTQE	780
Qy	780	QYSEYTGATYRWARSHEDGSEKDDWQRCTEIVADLHFRYLDQFVPEKVRRELNK	839
Db	781	QYSEYTGATYRWARSHEDGSEKDDWQRCTEIVADLHFRYLDQFVPEKVRRELNK	840



```

QY 840 AYCGLRPGVPSNLSAVATGNGCGAFGADARLKALIOILAAARERDVVYFTGDSL 899
Db 841 AYCGLRPGVPSNLSAVATGNGCGAFGADARLKALIOILAAARERDVVYFTGDSL 900
QY 900 MRDIYSMTFLTERKLDVGKYLRLRYNEECNCTGPGDIKLYPFYIHAVESAE 959
Db 901 MRDIYSMTFLTERKLDVGKYLRLRYNEECNCTGPGDIKLYPFYIHAVESAE 960
QY 960 D 960
Db 961 D 961

RESULT 3
Q8CB72 PRELIMINARY; PRT; 920 AA.
AC Q8CB72;
DT 01-MAR-2003 (TReMBLrel. 23, Created)
DT 01-MAR-2003 (TReMBLrel. 23, Last sequence update)
DT 01-JUN-2003 (TReMBLrel. 24, Last annotation update)
DE Poly.
GN PARG.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Bone;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium,
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RL Nature 420:563-573 (2002).
DR EMBL; AK036656; BAC29519.1; -.
DR MGD; MGI:1347094; Parg.
DR InterPro; IPR007724; PARG.
DR Pfam; PF05028; PARG; 1.
SQ SEQUENCE 920 AA; 103799 MW; 379E779E121B09A1 CRC64;

Query Match 94.3%; Score 4869.5; DB 11; Length 920;
Best Local Similarity 99.5%; Pred. No. 0;
Matches 945; Conservative 2; Mismatches 2; Indels 1; Gaps 1;

QY 1 MSAGPGWEPCTK-ARWGAAGTSAPTASDSRSFPGQRRLVDPKADPVQFRVPSSPACVS 59
Db 1 MSAGPGWEPCTKPRWGAAGTSAPTASDSRSFPGQRRLVDPKADPVQFRVPSSPACVS 60
QY 60 GRAGPHRGNATSFVFKQKTTTWMDTGPKTASSEKENNTRIDSMSSVQKDNFYPHK 119
Db 61 GRAGPHRGNATSFVFKQKTTTWMDTGPKTASSEKENNTRIDSMSSVQKDNFYPHK 120
QY 120 VEKLENPQLNLDKSPTEKSSQYLNQOQTASVCKWQNEKGHAEOLLASEPPAGTLPKQL 179
Db 121 VEKLENPQLNLDKSPTEKSSQYLNQOQTASVCKWQNEKGHAEOLLASEPPAGTLPKQL 180
QY 180 SNANIGOSPHTDHDSDTHDEDRNQOFLPIKLANTKPTVGDQARSNCKSGSRQSVK 239
Db 181 SNANIGOSPHTDHDSDTHDEDRNQOFLPIKLANTKPTVGDQARSNCKSGSRQSVK 240
QY 240 DCTCQOEVDVLPESPLSDVGAEDIGTPKNDKLTGQSSSLGDSPPFEKESPESPMD 299
Db 241 DCTCQOEVDVLPESPLSDVGAEDIGTPKNDKLTGQSSSLGDSPPFEKESPESPMD 300
QY 300 VDNERNSCQSEADEETSPVDEQDRSSQTANKLSSCOAREADGLRKRLTKGSEVRL 359
Db 301 VDNERNSCQSEADEETSPVDEQDRSSQTANKLSSCOAREADGLRKRLTKGSEVRL 360
QY 360 HFQFEGENNACTSDLNAPKSGNSSSLNVECRSSKHQGRKDSKITDHFMRISKSDRRKEQ 419
Db 361 HFQFEGENNACTSDLNAPKSGNSSSLNVECRSSKHQGRKDSKITDHFMRISKSDRRKEQ 420

```

```

QY 420 CEVRHQRTERKIPKYPNLPPEKKWLGTPIDEMKMPRCGHLPSLRPSASHTVTRVD 479
Db 421 CEVRHQRTERKIPKYPNLPPEKKWLGTPIDEMKMPRCGHLPSLRPSASHTVTRVD 480
QY 480 LIRAGEVPKPPFTHYKDLWDNKHVMPQSEONLYPVEDENGERTAGSRWELIQTALLNKF 539
Db 481 LIRAGEVPKPPFTHYKDLWDNKHVMPQSEONLYPVEDENGERTAGSRWELIQTALLNKF 540
QY 540 TRPONLKDAILKYNVAYSKKWDFTALVDVDFWDLSEAEBAQHLYQSILPDMVKIALCLPNI 599
Db 541 TRPONLKDAILKYNVAYSKKWDFTALVDVDFWDLSEAEBAQHLYQSILPDMVKIALCLPNI 600
QY 600 CTQPIPLLKQKXNHSVTMSQEQIASILLANAFCTFPRRNAMKSEYSSYPDINFNRLFEG 659
Db 601 CTQPIPLLKQKXNHSVTMSQEQIASILLANAFCTFPRRNAMKSEYSSYPDINFNRLFEG 660
QY 660 RSSRKPEKLTLCYFRRVTEKKPTGLVFTTQSLSEDFPEWERCCEKPLTRLHVYEGTIE 719
Db 661 RSSRKPEKLTLCYFRRVTEKKPTGLVFTTQSLSEDFPEWERCCEKPLTRLHVYEGTIE 720
QY 720 GNGRGMLOVDFAVRVGGVTCAGLVQEEIRFLINPELIVSRILFTEVLDHNECLIIITGE 779
Db 721 GNGRGMLOVDFAVRVGGVTCAGLVQEEIRFLINPELIVSRILFTEVLDHNECLIIITGE 780
QY 780 QYSEYTGVAETRYMARSHEDGSEKDDQRRCTEIVAIDALHFRRYLDQFVPEKVRRELNK 839
Db 781 QYSEYTGVAETRYMARSHEDGSEKDDQRRCTEIVAIDALHFRRYLDQFVPEKVRRELNK 840
QY 840 AYCGLRPGVPSNLSAVATGNGCGAFGADARLKALIOILAAARERDVVYFTGDSL 899
Db 841 AYCGLRPGVPSNLSAVATGNGCGAFGADARLKALIOILAAARERDVVYFTGDSL 900
QY 900 MRDIYSMTFLTERKLDVGK 919
Db 901 MRDIYSMTFLTERKLDVGK 920

RESULT 4
Q8QYM2 PRELIMINARY; PRT; 972 AA.
AC Q8QYM2;
DT 01-MAY-2000 (TReMBLrel. 13, Created)
DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)
DT 01-JUN-2003 (TReMBLrel. 24, Last annotation update)
DE Poly(ADP-ribose) glycohydrolase.
GN PARG.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BUF; TISSUE=Colon;
RA Shimokawa T., Masutani M., Nagasawa S., Nozaki T., Ikota N., Araki S.,
RA Aoki Y., Nakagawa H., Sugimura T.;
RT "Isolation and cloning of rat poly(ADP-ribose) glycohydrolase
RT (Parg).";
RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB019366; BAA87901.1; -.
DR GO; GO:0016787; F:hydrolase activity; IEA.
DR InterPro; IPR007724; PARG.
DR Pfam; PF05028; PARG; 1.
DR Hydrolase.
SQ SEQUENCE 972 AA; 109002 MW; 3DED7885684F59E9 CRC64;

Query Match 92.0%; Score 4749; DB 11; Length 972;
Best Local Similarity 92.0%; Pred. No. 0;
Matches 894; Conservative 26; Mismatches 48; Indels 4; Gaps 2;

QY 1 MSAGPGWEPCTK-ARWGAAGTSAPTASDSRSFPGQRRLVDPKADPVQFRVPSSPACVS 59
Db 1 MSAGPGWEPCTKPRWGAAGTSAPTASDSRSFPGQRRLVDPKADPVQFRVPSSPACVS 60

```



772 CLIIITGEQSYTGYAETRYRWSHEDSGKDDWQRCTEIVADALHFRYLDQFVPE 831  
780 CLIIITGEQSYTGYAETRYRWSHEDSGKDDWQRCTEIVADALHFRYLDQFVPE 839  
832 KVRRELKAYCGFLRPGVSENLAVATGCGGFGDARLAKALIQLIAAAAAARDVY 891  
840 KVRRELKAYCGFLRPGVSENLAVATGCGGFGDARLAKALIQLIAAAAAARDVY 899  
892 FTFGDSELMDIYSMTFTLTKRLDVGKVKYKLLRYNEECRCNSTPGPDIKLYPIIYHA 951  
900 FTFGDSELMDIYSMTFTLTKRLDVGKVKYKLLRYNEECRCNSTPGPDIKLYPIIYHA 959  
952 VESSAETTDMPGQKAGT 968  
960 VESCAETADHSGQRTGT 976

RESULT 6  
Q7Z742 ID Q7Z742 PRELIMINARY; PRT; 976 AA.  
AC Q7Z742; MEDLINE=22388257; PubMed=12477932;  
DT 01-OCT-2003 (TrEMBLrel. 25, Created)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Hypothetical protein.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Testis;  
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.F., Bhat N.K.,  
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Haieh F.,  
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.B.,  
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,  
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaby S.J.,  
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,  
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,  
RA Krzywinski M.I., Stalska U., Smailus D.E., Schnerch A., Schein J.B.,  
RA Jones S.J., Marra M.A.;  
RT "Generation and initial analysis of more than 15,000 full-length human  
RT and mouse cDNA sequences.";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Testis;  
RA Strausberg R.;  
RL Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.  
DR EMBL; BC052966; AAH52966.1; --  
KW Hypothetical protein.  
SQ SEQUENCE 976 AA; 111090 MW; 47C8EE1826CA474A CRC64;

Query Match 86.2%; Score 4451; DB 4; Length 976;  
Best Local Similarity 86.5%; Pred. No. 0;  
Matches 845; Conservative 49; Mismatches 73; Indels 10; Gaps 6;  
Qy 1 MSAGFGEPCPK-ARWGAAGTSAFTASDSRSPGQRRLVDPKADPQVFRPPSPACVS 59  
Db 1 NNAGFGEPCPKRPRWGAATTPSPAASDARSFPRQRRLVDPKADPQVFRPPSPACVP 59  
Qy 60 GRAGPHRGNATSFVFKQXTITWMDTKGPKTAESE---SKENNTTRIDSMSSVQKDNFY 116  
Db 60 GRAGPHRGNATSLVFKQXTITWMDTKGPKTAESLDSKENNTTRIDSMSSVQKDNFY 119

Qy 117 PHKVEKLENVPLNDKSPTEKSSOYLNQOQTASVCKWQNEKHAELQLLASPPACTPLP 176  
Db 120 QNVKLENVPLNDKSPTEKSTQTLNQHQTAACKWQNEKHTQLLESFQTVLVP 179  
Qy 177 KQLSNANTGQSPHTDDHSDTDHEEDRDNQOFLTPIKLANTKPTVGDQAR---SNCKCSG 233  
Db 180 EQFSNANIDRSFQNDHSDTDSEENRDNQOFLTTVKLANAKQTTEDQAREAKSHOKCSK 239  
Qy 234 SRSQVKDCTGCOQBEVDVLPSPSLDVGNAEDTGTGPNKDKLTGOSLSLGDSPPEKSE 293  
Db 240 SCDPDCASCQOQDIDVVPESPLSDVGSDEVTGTPNDKDKLTGQSCUGNPPPEKSE 299  
Qy 294 PESPMVDNRSNCSQDSEADEETSPVFDQDDP-RSSQTANKLSSCOAREADGRLRYLT 352  
Db 300 PESPMVDNRSNCSQDSEADEETSPGFDQEDGSSSQTANKPSRFQARDADIEFRKYST 359  
Qy 353 KGSVRLHFQFE-GENNAGTSDLNKPSNSSLNVECSRSSQKHGKRDSDKITDHFWRISK 411  
Db 360 KGSVRLHFQFEGESRTGMDLNKLPNIGSSLNVECSRSSQKHGKRDSDKITDHFWRISK 419  
Qy 412 SDRRKEQCEVHQRTKIPKYPNLPPEKKWLGTPTEEMRMPCRCGHLPLSPSPAS 471  
Db 420 AEDRRKEQWETKHQRTKIPKYPNLPPEKKWLGTPTEEMRMPCRCGHLPLSPSPAS 479  
Qy 472 HTVTVRDILLRAGEVPKPPPTHYKDLNDKNKVMPCSEQLNLYPVEDENGERTAGSRWELI 531  
Db 480 HTVTIRVDILLRAGEVPKPPPTHYKDLNDKNKVMPCSEQLNLYPVEDENGERTAGSRWELI 539  
Qy 532 QTALLNKFTRPQNLKDAIILKNVAYSKKWDFLTALVDFWDKVLEEAQAHLIYQSILDMVK 591  
Db 540 QTALLNKFTRPQNLKDAIILKNVAYSKKWDFLTALVDFWDKVLEEAQAHLIYQSILDMVK 599  
Qy 592 IALCLPNICTQPIPLKQKNHSTVMSQSIASLANAFCTEPRRNAMKSSYSSYPDI 651  
Db 600 IALCLPNICTQPIPLKQKNHSTVMSQSIASLANAFCTEPRRNAMKSSYSSYPDI 659  
Qy 652 NFNRLEGGSSSRKPEKLTLCVFRVTEKPTGLVTFTRQSLDEDPFEMERCEKPLTRLH 711  
Db 660 NFNRLEGGSSSRKPEKLTLCVFRVTEKPTGLVTFTRQSLDEDPFEMERCEKPLTRLH 719  
Qy 712 VTVEGTIEGNGRGLQVDFANFRVGGVGTGAGLVQBEIRFLINPELIVSLFTEVLDHNE 771  
Db 720 VTVEGTIEGNGRGLQVDFANFRVGGVGTGAGLVQBEIRFLINPELIVSLFTEVLDHNE 779  
Qy 772 CLIIITGEQSYTGYAETRYRWSHEDSGKDDWQRCTEIVADALHFRYLDQFVPE 831  
Db 780 CLIIITGEQSYTGYAETRYRWSHEDSGKDDWQRCTEIVADALHFRYLDQFVPE 839  
Qy 832 KVRRELKAYCGFLRPGVSENLAVATGCGGFGDARLAKALIQLIAAAAAARDVY 891  
Db 840 KVRRELKAYCGFLRPGVSENLAVATGCGGFGDARLAKALIQLIAAAAAARDVY 899  
Qy 892 FTFGDSELMDIYSMTFTLTKRLDVGKVKYKLLRYNEECRCNSTPGPDIKLYPIIYHA 951  
Db 900 FTFGDSELMDIYSMTFTLTKRLDVGKVKYKLLRYNEECRCNSTPGPDIKLYPIIYHA 959  
952 VESSAETTDMPGQKAGT 968  
960 VESCAETADHSGQRTGT 976

RESULT 7  
Q9Y4W7 ID Q9Y4W7 PRELIMINARY; PRT; 976 AA.  
AC Q9Y4W7;  
DT 01-NOV-1999 (TrEMBLrel. 12, Created)  
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
DE Poly (ADP-ribose) glycohydrolase.  
GN HPARG.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC	Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
NCBI_TaxID=9606;	
[1]	
SEQUENCE FROM N.A.	
MDL=93380098; Pubmed=10449915;	
Ane J.C., Apion F., Jacobson E.L., Jacobson M.K.;	
"Assignment of the poly(ADP-ribose) glycohydrolase gene (PARG) to	
human chromosome 10q11.23 and mouse chromosome 14B by in situ	
hybridization."	
Cytogenet. Cell Genet. 85:269-270 (1999).	
EMBL; AF005043; AAB61614.1; -.	
Genew; HGNC:8605; PARG.	
GO: 0005737; C:cytoplasm; TAS.	
GO: 0004649; F:poly(ADP-ribose) glycohydrolase activity; TAS.	
InezFro; IPR007724; PARG.	
Pfam; PF05028; PARG; 1.	
Hydrolase.	
SEQUENCE 976 AA; 110903 MW; 42A0F51466FCE7DC CRC64;	
Query Match	85.1%; Score 4395; DB 4; Length 976;
Best Local Similarity	85.7%; Pred. No. 0;
Matches	83; Conservative 78; Indels 10; Gaps 6;
QY	1 MSAGFGNEPCTKA-RWGAAGTSAPTASDSRSPGQRRLVDPKAPVQFRPPSPACVS 59
DB	1 MNAGFGCEPCTKATRWGAA--TTSPAASDARSFPQRRLVDPKDAHVQFRPPSPACVP 59
QY	60 GRAGPHRGNATSFVFKQKTIITWMDTKGPKTAES--SKENNNTRIDSMSSVQKONFY 116
DB	60 GQAGQHRGASLSLVFKQKTIISWMDTKGTAESLSLDSKNNNTRIESMSSVQKONFY 119
QY	117 PHVKEKLENVQLNDKSPTEKSGOYLNQOQTASVCKQWNGKQKARQALLASEPPAGTLP 176
DB	120 QHNVEKLVNVSQSLDKSLTEKSTQYLNQHQTAAWCKWNGKTEQLLESEPOVTILVP 179
QY	177 KQLSNANIGSPHTDDHSDTHEEDRDNQOFTPIKLANTRPTVGDGAR--SNCKCSG 233
DB	180 EQFSNANIDRSPQNDHSDTUSEENRDNQOFTVTVKLANAQTTEDFAHRAKSHQKSK 239
QY	234 SRQSVKDCGCOQEEVDVLPSPSLSDVGAEDTGTGPKNDKLTGQESSLGSPPPEKSE 293
DB	240 SCHPEDCASQOQEDIDVVPKSPSLSDVSGEDVGTGSKNDKLIQGESCLGNSPPPEKSE 299
QY	294 PESPMVDVNSRNSQDSEADETSPVFEQDD-RSSQTANKLSSCOAREADGLRKRYLT 352
DB	300 PESPMVDVNSKNSQDSEADETSPGFQEQDSSQTANKFSRQADADLEFRKRIST 359
QY	353 KGSEVRLHQPPE-GRNNAGTSDLNAPKSGNSSLNVECRSSQHGKGRDSKITDHPMRISK 411
DB	360 KGEVRLHQPFGESRPTGMDLNAKLPNISSLNVECRNSQHGKQKSDKITDHLMLPLP 419
QY	412 SEDRRKEQCEVHRQTERKIPKYPINLPPEKKWLTPIEHRMKPRGCIHLPSPSPNS 471
DB	420 AEDRNKEQWETKHQTERKIPKYPVPHLSPDKKWLGTPIEHRMRPRGIRLPLRPSAN 479
QY	472 HTVTYRVOLLRAGEVPKPEPTHYKOLMDNKHVKMPCSEQLYPVEDENGERTAGSRWELI 531
DB	480 HTVTIKVDLLRAGEVPKPFPTHYKOLMDNKHVKMPCSEQLYPVEDENGERTAGSRWELI 539
QY	532 QTALLNKFTRPQMLKDAILKYNVAYSKKWDFTALVDFWDKVLFEAAQHLYSILPDMVK 591
DB	540 QTALLNKFTRPQMLKDAILKYNVAYSKKWDFTALIDFWDKVLFEAAQHLYSILPDMVK 599
QY	592 IALCLPNICTQPIPLIKQKNHVSVMSEQIASILANAFCTFPRNKKMKSEYSSVPI 651
DB	600 IALCLPNICTQPIPLIKQKNHVSITMSQEQIASILANAFCTFPRNKKMKSEYSSVPI 659
QY	652 NFNRLEFGSSRRPEKLTLCVFRVTEKKPTGLVTFRQSLDPEPWERCEKPLTRLH 711
DB	660 NFNRLEFGSSRRPEKLTLCVFRVTEKKPTGLVTFRQSLDPEPWERCEKPLTRLH 719
QY	712 VTYEGTIEGNRGMQVDFANRRFVGGSVTGAGIVQSEIRPLINPELIVSLFTEVLDDHNE 771

Db	720	VYEGTIENGQGMQLQVDFANFVGGVTSAGLVQEEIRFLINPELLISRLFTVELDHNE	779			
Qy	772	CLIIITGEQYSEYTGAEIYRWARSHEDSEKDDQWRRCCTEIVADALHPRRYLDQFVPE	831			
Db	780	CLIIITGEQYSEYTGAEIYRWSRSHEDGSRDDCERRCTEIVADALHPRRYLDQFVPE	839			
Qy	832	KVRELNKAYCGFLRPGVPSENLSAVATGNWGCAGFGGDARLKALICILAAAAAERDVVY	891			
Db	840	KVRELNKAYCGFLRPGVPSENLSAVATGNWGCAGFGGDARLKALICILAAAAAERDVVY	899			
Qy	892	FTGDSSELMRDIIYSMETETETERKLDGVKVKLLRYNEECRCNCSTPGPDIKLYPIYIHA	951			
Db	900	FTGDSSELMRDIIYSMHILFETERKLTGVDVYKLLRYNEECRCNCSTPGPDIKLYPIYIHA	959			
Qy	952	VESSAETTDMPGOKAGT 968				
Db	960	VESCAETADHSGORTGT 976				
RESULT 8						
ID	002776	PRELIMINARY; PRT; 977 AA.				
AC	002776;					
DT	01-JUL-1997	(TREMBlrel. 04, Created)				
DT	01-JUL-1997	(TREMBlrel. 04, Last sequence update)				
DT	01-JUN-2003	(TREMBlrel. 24, Last annotation update)				
DE	Poly(ADP-ribose) glycohydrolase.					
GN	BPARG.					
OS	Bos taurus (Bovine).					
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;					
OC	Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidea;					
OC	Bovidae; Bovinae; Bos.					
OX	NCBI_TaxID=9913;					
RN	[1]					
RP	SEQUENCE FROM N.A.					
RX	MEDLINE=97277328; PubMed=9115250;					
RA	Lin W., Ame J.C., Abouli-Ela N., Jacobson E.L., Jacobson M.K.;					
RA	"Isolation and characterization of the cDNA encoding bovine poly(ADP-					
RT	ribose) glycohydrolase."					
RL	J. Biol. Chem. 272:11895-11901(1997).					
DR	EMBL; U78975; AAB53370.1; -. Hydrolase activity; IEA.					
DR	GO; GO:0016787; F:Hydrolase activity; IEA.					
DR	InterPro; IPR007724; PARG.					
DR	Pfam; PF05028; PARG; 1.					
KW	Hydrolase.					
SQ	SEQUENCE 977 AA; 110837 MW; 87D2100F979DF377 CRC64;					
Query Match						
Best Local Similarity 83.9%; Score 4332.5; DB 6; Length 977;						
Matches 820; Conservative 57; Mismatches 90; Indels 9; Gaps 5						
Qy	1	MSAGPGWPCPTK-ARWGAAGTSAPTASDSRSPFQRRVLDPKAPVQFRVPPSPACVS	59			
Db	1	MSAGPGWPCPTKRPWDAAATSPPAASDARSPFQRRVLDPKAPVQFRVPPSSGCGAL	60			
Qy	60	GRAGPHRGNATSFVEFKTITTMWDTKGPKTAESE---SKENNTRIISMSSVQKNFY	116			
Db	61	GRAGPHRGNATSLVEFKTITTSWMDTKGIVTESLSLHSENNNTRSESMSSVQKNFY	120			
Qy	117	PHKVEKLENVPLNLDKSPTEKSSQYLNQQTASVCKWQNGKHAQQLLASEPPAGTPLP	176			
Db	121	QHNMEKLENVSLGFDKSPVEKGTQYLNQQTAAWCKWQNGKHAQQLLASEPPAVTLVP	180			
Qy	177	KQLSNANIGOSPHTDDSHSDTHDEDRDQQFLTPILKANTKPTVGDGQ---ARSNCKSG	233			
Db	181	EQFSNANVDQSPKDDHSDHTNSESRDQQFLTHVKLANAKQTMEDQEGREARSHQCKG	240			
Qy	234	SRQSKVDCITGCQCEEVVDVLPESFLSDVGAEDIGTGPKNDKNLTGQESSLGDSPPPEKSE	293			
Db	241	ACHPAEACAGCQCEETDVVSEFSLDTGSEDVGTGLKANRLNRQESSLGDSPPPEKSE	300			
Qy	294	PESPMDVNSRNSQDSSEADETSPVFEQDDRSS-QTANKLSSCQAREADGLARKYLT	352			

301 PESPMVDVNSKNSCODSEADSETSPGDEQEDSSAQANKPESRFPQPREADTELKRGSA 360  
353 KGSEVRLHFQFE-GENNAGTSDLNKAPGNSSSINVECRSSKQHGKRSKIDTHFWRLSK 411  
361 KGSEIRLHFQFEGESRAGDNDVNAKRGSTSLNVECRSSKQHGKRSKIDTHFWRVPK 420  
412 SDRKKEQCEVHQTERKIPKYPINLPPEKKWLTGTPIEEMKMPKRCGHIPLSPRPSAS 471  
421 AEDKKEQCEMHQTERKIPKYPINLPPEKKWLTGTPIEEMKMPKRCGHIPLSPRPSAN 480  
472 HVTIVRVLLRAGEVPPPTTHYKOLMDNKHVMPQSEONLYVPEDENGERTAGRWELI 531  
481 HVTIVRVLLRAGEVPPPTTHYKOLMDNKHVMPQSEONLYVPEDENGERTAGRWELI 540  
532 QPALNKFTRPNQKDAILKYNVAYSKKDFVALDVFWDKVLAEAAQHLYSILPDMVK 591  
541 QPALNKFTRPNQKDAILKYNVAYSKKDFVALDVFWDKVLAEAAQHLYSILPDMVK 600  
592 IALCLPNICTOPTILKOKMNSVTSQEQIASLLANAFCTPPRRNAKMSSEYSPDI 651  
601 IALCLPNICTOPTILKOKMNSVTSQEQIASLLANAFCTPPRRNAKMSSEYSPDI 660  
652 NFNRLEPGRSSRKPETKLCFCYFRVTEKTKPTGLVTFTRQSLDPPWERCEKPLTRLH 711  
661 NFNRLEPGRSSRKPETKLCFCYFRVTEKTKPTGLVTFTRQSLDPPWERCEKPLTRLH 720  
712 VTYEGTIEGNGOMLQVDFANRVGGVTCAGLVGEIRFLINPELIVSRLFEVLHDNE 771  
721 VTYEGTIEGNGOMLQVDFANRVGGVTCAGLVGEIRFLINPELIVSRLFEVLHDNE 780  
772 CLIIITGEQYSEVTGAETVRSWASHEDGSEKDDWRRCTEIVADLHFRRLDQFVPE 831  
781 CLIIITGEQYSEVTGAETVRSWASHEDGSEKDDWRRCTEIVADLHFRRLDQFVPE 840  
832 KVRRELKAYCGFLRPGVPSNLSAVATNGWCGAFGGDARLKALIQILAAAAERDVY 891  
841 KVRRELKAYCGFLRPGVPSNLSAVATNGWCGAFGGDARLKALIQILAAAAERDVY 900  
892 FTFGDSLMRDVISMHTFLTERKLDVGVKYLKLLRYNECRNCSTPGPDILKYPPIYA 951  
901 FTFGDSLMRDVISMHTFLTERKLDVGVKYLKLLRYNECRNCSTPGPDILKYPPIYA 960  
952 VESSAETDMPGQKAG 967  
961 VESCTQTTNPQORTG 976  
RESULT 9  
O46043 PRELIMINARY; PRT; 768 AA.  
AC O46043  
DT 01-JUN-1998 (TREMELrel. 06, Created)  
DT 01-MAY-1999 (TREMELrel. 10, Last sequence update)  
DT 01-JUN-2003 (TREMELrel. 24, Last annotation update)  
DE PARG protein.  
GN PARG OR SG:114E2.1 OR CG2864.  
OS Drosophila melanogaster (fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7227;  
RN [1]  
FP SEQUENCE FROM N.A.  
RC STRAIN=Berkeley;  
RX MEDLINE=20196006; PubMed=10731132;  
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,  
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
RA Sutton G.G., Wortman J.R., Vandeil M.D., Zhang Q., Chen L.X.,  
RA Brannon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,  
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
RA Abril J.F., Aghayani A., An H.-J., Andrews-Pfankuch C., Baldwin D.,  
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,

RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,  
RA Burtis K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,  
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,  
RA Fostler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasner K.,  
RA Glöckler A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,  
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissenbach J.,  
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
RA Zheng X.H., Zhong F.N., Zhou W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
RT "The genome sequence of Drosophila melanogaster.";  
RL Science 287:2185-2195(2000).  
RN [2]  
RP SEQUENCE FROM N.A.  
RP Ame J.-C., Jacobson M.K.;  
RT "Isolation and characterization of the cDNA encoding Drosophila  
RT poly (ADP-ribose) glycohydrolase.";  
RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE OF 46-768 FROM N.A.  
RA Murphy L., Harris D., Barrell B.;  
RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.  
RN [4]  
RP SEQUENCE OF 46-768 FROM N.A.  
RA Benos P.;  
RL Submitted (DEC-1997) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AE003428; AAF45886.1; -;  
DR EMBL: AF079556; AAC28734.1; -;  
DR EMBL: Z98254; CAB10913.1; -;  
DR FlyBase; FBgn0023216; Parg.  
DR GO; GO:0016787; F:hydrolase activity; IEA.  
DR InterPro; IPR007724; PARG.  
DR Pfam; PF05028; PARG; 1.  
DR Hydrolase.  
SQ SEQUENCE 768 AA; 86288 MW; 2E620C42E643FC04 CRC64;  
Query Match 20.8%; Score 1074.5; DB 5; Length 768;  
Best Local Similarity 41.7%; Pred. No. 3.5e-71;  
Matches 244; Conservative 96; Mismatches 194; Indels 51; Gaps 15;  
Qy 408 RISKSDRRKEQCEVHQTERKIPKYPINLPPE--EKKWLTGTPIEEMKMPKRCGHIPLS 465  
Db 45 RMSKSPDGGIGSEITREE-----PENLANSLLDSSWRGVSMETAIHR-NRQPFLEN 93  
Qy 466 LRP-SASHTVTVRVLLRAGEVPPPTTHYKOLMDNKHVMPQSEONLYVPEDENGERTAG 524  
Db 94 LPPVTAGNLHRVMYQVPIRETPPRPKYSPGK--WDSEHVLPCAPESKYPRENDPGSTTI 151  
Qy 525 GSRWELIQTALLNKFTPQNLKDAILKYNVAYSKKDFVALDVFWDKVLAEAAQHLYSQS 584  
Db 152 DFRWEMIERALLOPIKTCCEBLOAAIISYNTYTDQMFHRLHQLLDDEELDESETRVFED 211  
Qy 585 ILPMVKIALCLPNICTOPTILKOKMNSVTSQEQIASLLANAFCTPPRRNAKMSSEYSPDI 643  
Db 212 LLPRIIRLRLPLDIQSPFELKHHKNVSNLSUSQOQISCLLANAFCTPPRRNLTKRS 271

Qy	644	EYSSYPDINFRNLRPEGRSSRKPKEKLTLCYFVRV--TEKK-----PTGLVTFTRQS--LE	695
Db	272	EYSFPDINFRNLYQSTGPAVLEKLCIMHYFRRVCPTRDASNVPTGVTFFVRSGLPE	331
Qy	696	DPFWRCEKPL--TRLHVTEGIEGNRGMLOVDFAFRFVGGVTCAGLVQVEIRFLI	753
Db	332	HLIDWSQAAPLGDVPLHVDAGETIEDGIGLLQVDFANKYLGGVGLGHGCVQVEIRFVI	391
Qy	754	NPELIVSRLEFTEVLNDHNECLITGTEQYSEYTGVAETVYMARSHEDGSEKDDWQRCITEI	813
Db	392	CPELLVGLKFTCELRPFALVWLGAERYSNYTGAGSPFWSGNFEDSTPRDSSGRQTAI	451
Qy	814	VAIDALHFRYLDQFVPEKVRRELNKAYCGFLR-----PGVPSNLGAVATGNWCGAF	867
Db	452	VAIDALHFAQSHHQVREDLMERELNKAYIGFVHMVTPPPG-----VATGNWCGAF	503
Qy	868	GGDARLKALIQILAAAAERDVVYTFGDSLEMRDIYSMHTFLTERKLDVGVYKLLRY	927
Db	504	GGDSYLKALLQLMVCAQLGRPLAYYTFGNVFRDDFHEMWLLFRNDGTTVQQLWS--ILRS	562
Qy	928	YNEECRNCSTPGP-----DIKLYPIYHAVESSAETDMPGQKA	966
Db	563	YSRLIKESKSEPRENKASKKLYDFI---KEELKKVRDVPGEA	604
RESULT 10			
ID	Q960N8	PRELIMINARY;	PRT; 723 AA.
AC	Q960N8;		
DC	01-DEC-2001 (TREMELrel. 19, Created)		
DT	01-DEC-2001 (TREMELrel. 19, Last sequence update)		
DT	01-JUN-2003 (TREMELrel. 24, Last annotation update)		
DE	LD42380P.		
GN	PARG OR EG:11482.1 OR CG2864.		
OS	Drosophila melanogaster (Fruit fly).		
OC	Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;		
OC	Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;		
OC	Ephydroidea; Drosophilidae; Drosophila.		
OX	NCBI_TaxID=7227;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN-Berkeley;		
RA	Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,		
RA	Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,		
RA	Gonzalez M., Guatin H., Li P., Liao G., Miranda A., Mungali C.J.,		
RA	Nunco J., Pacleb J., Paragas V., Park S., Phouanavong S., Wan K.,		
RA	Yu C., Lewis S.E., Rubin G.M., Celniker S.		
RL	Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.		
DR	EMBL; AF051955; AAK93379.1; -		
DR	FlyBase; FBgn0023216; Parg.		
DR	InterPro; IPR007724; PARG.		
DR	Pfam; PF05028; PARG; 1.		
SQ	SEQUENCE 723 AA; 81109 MW; 90A5BA59775D611C CRC64;		
Query Match			
Best Local Similarity 20.7%; Score 1069.5; DB 5; Length 723;			
Matches 243; Conservative 96; Mismatches 194; Indels 51; Gaps 15;			
Qy	409	IKSDERRKEOCVRHQRTERKIPVNLPP--EKKWLGTPIERMKMPCRGHLPGL	466
Db	1	MSKSPGGGISEITEE-----PENLNSLDDSNRGVSMIAHR-NRQPFLENL	49
Qy	467	RP-SASHVTVRVLLRAGEVPPFPPTHYKOLMDNKHVMPCSQNLVPEVDENGERTAG	525
Db	50	PPVTAGNLHRVMYQLPIRETPPPPYKSPGK--WDESHRPLCAPESKYPRENPDGTTID	107
Qy	526	SRWELQTALLNKFTPNKLDAILKXNAVYSKKWDFALVDFWVKVLEAEAHLYQSI	585
Db	108	FRWEMIERALLOPIKTCCEELQAAIISYNTYRDQMHFRALHQLLDELDESETRVFFDL	167
Qy	586	LPDMVKIALCLPNICTOPIPLLKQKMHSVTMSQEQIASLANAFCTTFRPNA--RMKSE	644
Db			
Qy	168	LPRIIRLALRLPDLIQSPVPLLKHKNASLSLSQQQISCLIANAFCTFPRRNTLRKKE	227
Qy	645	YSSYPDINFRNLRPEGRSSRKPKEKLTLCYFVRV--TEKK-----PTGLVTFTRQS--LE	696
Db	228	EYSFPDINFRNLYQSTGPAVLEKLCIMHYFRRVCPTRDASNVPTGVTFFVRSGLPE	287
Qy	697	DPFWRCEKPL--TRLHVTEGIEGNRGMLOVDFAFRFVGGVTCAGLVQVEIRFLI	754
Db	288	HLIDWSQAAPLGDVPLHVDAGETIEDGIGLLQVDFANKYLGGVGLGHGCVQVEIRFVI	347
Qy	755	NPELIVSRLEFTEVLNDHNECLITGTEQYSEYTGVAETVYMARSHEDGSEKDDWQRCITEI	814
Db	348	PPELLVGLKFTCELRPFALVWLGAERYSNYTGAGSPFWSGNFEDSTPRDSSGRQTAI	407
Qy	815	VAIDALHFRYLDQFVPEKVRRELNKAYCGFLR-----PGVPSNLGAVATGNWCGAF	868
Db	408	VAIDALHFAQSHHQVREDLMERELNKAYIGFVHMVTPPPG-----VATGNWCGAF	459
Qy	869	GGDARLKALIQILAAAAERDVVYTFGDSLEMRDIYSMHTFLTERKLDVGVYKLLRY	928
Db	460	GGDSYLKALLQLMVCAQLGRPLAYYTFGNVFRDDFHEMWLLFRNDGTTVQQLWS--ILRS	518
Qy	929	NEECRNCSTPGP-----DIKLYPIYHAVESSAETDMPGQKA	966
Db	519	SRLIKESKSEPRENKASKKLYDFI---KEELKKVRDVPGEA	559
RESULT 11			
ID	Q98KB3	PRELIMINARY;	PRT; 548 AA.
AC	Q98KB3;		
DC	01-MAY-2000 (TREMELrel. 13, Created)		
DT	01-JUN-2002 (TREMELrel. 21, Last sequence update)		
DT	01-OCT-2003 (TREMELrel. 25, Last annotation update)		
DE	Putative poly (ADP-ribose) glycohydrolase.		
GN	AT2G31870 OR TEJ.		
OS	Arabidopsis thaliana (Mouse-ear cress).		
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;		
OC	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;		
OC	eurosid II; Brassicales; Brassicaceae; Arabidopsis.		
OX	NCBI_TaxID=3702;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=cv. Columbia;		
RA	Lin X., Kaul S., Shea T.P., Fujii C.Y., Shen M., VanAken S.E.,		
RA	Barnstead M.E., Mason T.M., Bowman C.L., Renning C.M., Benito M.-I.,		
RA	Carrera A.J., Creasy T.H., Suel C.R., Town C.D., Nierman W.C.,		
RA	Fraser C.M., Venter J.C.;		
RL	Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.		
DR	EMBL; AF394690; AAK72256.1; -		
DR	PIR; B84726; B84726.		
DR	GO; GO:0016787; F:hydrolase activity; IEA.		
DR	InterPro; IPR007724; PARG.		
DR	Pfam; PF05028; PARG; 1.		
KW	Hydrolase.		
SQ	SEQUENCE 548 AA; 62169 MW; F1A79FDA157C3329 CRC64;		
Query Match			
Best Local Similarity 11.0%; Score 568; DB 10; Length 548;			
Matches 147; Conservative 61; Mismatches 139; Indels 104; Gaps 11;			
Qy	569	WDKVLLEAEAHLYQSIPLDMVKIALCLPNI-----CTQPIPLKQKMHNS	614
Db	79	FELIDEXKSKWFEIIPALASLLQFFSLELVHFNQADNVSGIKTGLRLNSQAGI	138
Qy	615	VTMSQEQIASLANAFCTTFRPNAKMKSEYSDPDINRLEFE-----GRSSRKEPKLKT	670

Db 139 VFLSOLIGALLACFFCLFPDDNRGAK-----HLFVINFDHLFASLYTSYSQSQSKIRC 194  
QY 671 LFCYFERRVTEKPTGLVFTFTR--SLEDPPEWERCEKPLTRLHVITYEGTIEGNGRGLQV 728  
Db 195 IMHYFERFCSCVPIGIVSFERKITAPADAFKSDVSLCAFKVHSFGLIEDOPDPALEV 254  
QY 729 DFANFVGGVGTGAGLVQBEIRFLINPELIVSRLFTVLDHNECLIIITGTEOYSYTYGA 788  
Db 255 DFANKYLGGLSLRSCVQBEIRFLINPELIVSRLFTVLDHNECLIIITGTEOYSYTYGA 314  
QY 789 ETYRVARHEDSEKDDWRCTEIVADAL-----HFRYLDQVPEKVERELNKAYC 842  
Db 315 SSFRFAGEYIDKAMDPEKRRTRIIVADALCTPKMRHFKDIC-----LUREINKALC 367  
QY 843 GFL-----RPG-----VPSN 853  
Db 368 GFENCSKAWHQNI FMEGDNELQVNRGRDSGLLRTETTASHRTPLNDVEMREKPA 427  
QY 854 L-----SAVATGNWCGAFGGDARLALIIILAAAARDVV-YFTFGD 896  
Db 428 LIRDFVVEGVDNEDHDDGVATGNWCGVFGDPPELKATIQWLAASQTRRPFISYTFG- 486  
QY 897 SELMRDIYSMHTFLTERKLDVGKVKLLRY 927  
Db 487 VEALRNLDQVTKWILSHKWTVDGLMNMWLEY 517  
RESULT 12  
Q8VYAL PRELIMINARY; PRT; 522 AA.  
AC Q8VYAL;  
DT 01-MAR-2002 (TrEMBLrel. 20, Created)  
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
DE Putative poly (ADP-ribose) glycohydrolase.  
GN AT2G31870.  
OS Arabidopsis thaliana (Mouse-ear cross).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; Core eudicots; rosids;  
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.  
OX NCBI\_TaxID=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Nguyen M., Karlin-Neumann G., Southwick A., Lam B., Miranda M.,  
Palm C.J., Bowser L., Jones T., Banh J., Carninci P., Chen H.,  
Cheuk R., Chung M.K., Hayashizaki Y., Ishida J., Kamiya A., Kawai J.,  
Kim C., Lin J., Liu S.X., Narusaka M., Pham P.K., Sakano H.,  
Sakurai T., Satou M., Seki M., Shinn P., Yamada K., Shinozaki K.,  
Ecker J., Theologis A., Davis R.W.;  
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AY072330; AAL61937.1;  
DR GO; GO:0016787; F-hydrolase activity; IEA.  
DR InterPro; IPR007724; PARG.  
DR Pfam; PF05028; PARG; 1.  
KW Hydrolase.  
SQ SEQUENCE 522 AA; 58942 MW; 6FF98685D867CA15 CRC64;  
Query Match 10.2%; Score 525; DB 10; Length 522;  
Best Local Similarity 30.5%; Pred. No. 1.7e-30;  
Matches 132; Conservative 82; Mismatches 139; Indels 80; Gaps 11;  
QY 568 FWDKVLBEAQAHLQYQSILPMVXIALCLPNCITQPIPLKQKMH----- 613  
Db 80 FFDKKISRESANFGEVVPALCLLLQLPSMLEKH-----QKADHVLGDKVSKSLRLGP 135  
QY 614 ----SVTMSQEQIASLANAFCTPPRRNAKMKSESYDPDINRLE-----GRSRKP 665  
Db 136 QEAGIVLLSQELIAALLACSPFLCFPEVDRLK-----NLOGINFGSLFSPFMRHCKTQE 191  
QY 666 EKLKTLFCYFRVTEKPTGLVFTFTRQSL-----DPPE-----WERCEKPLRLHVITY 714  
Db 192 NKIKLTHYFGRICRMWPTGFVSFERKILPLEYHPHFVSYPKADSWANSVTPLCISIH 251

QY 715 EGTIEGNGRGLQVDFANFVGGVGTGAGLVQBEIRFLINPELIVSRLFTVLDHNECLII 774  
Db 252 SGATEDQCEALEVDFADEYFGGLTSLYDTQEEIRFVINPELIVSRLFTVLDHNECLII 311  
QY 775 ITGTEOYSYTYGAIRYVARHEDSEKDDWRCTEIVADALHFRYLDQVPEKVR 834  
Db 312 IVGVRFSGYTYGSPFOYAGDYTDNKLDIFRRKTRVIAIDAMP-DPGMGQYKLDALI 370  
QY 835 RELNKAYCGFLR-----PGVPSNLSA-----VA 858  
Db 371 REVNKAFSGYTHQCKYKINIDVHKDPEASSSHVSLTSDSASQVIESHRNCIDHEEKIGVA 430  
QY 859 TGNWCGGAFGGDARLALIIILAAAARDVV-YFTFGDSELMRDIYSMHTFLTERKLDV 917  
Db 431 TGNWCGVFGDPPELKIMLQWLAISQSGRPFMSYTYFG-LQALQNLQNVLENVALQEMTV 489  
QY 918 GKVKYKLLRYNE 930  
Db 490 GDLWKKLVESSE 502  
RESULT 13  
Q9NSL4 PRELIMINARY; PRT; 485 AA.  
AC Q9NSL4;  
DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
DE Hypothetical protein H23L24.5 (Poly ADP-ribose metabolism enzyme  
4)  
GN H23L24.5 OR PME-4.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;  
OC Rhabditidae; Peloderinae; Caenorhabditis.  
OX NCBI\_TaxID=6239;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Bristol N2;  
RX MEDLINE=99069613; PubMed=9851916;  
RA Waterston R.;  
RT "Genome sequence of the nematode C. elegans: a platform for  
investigating biology. The C. elegans Sequencing Consortium.";  
RL Science 282:2012-2018(1998).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Bristol N2;  
RA Abu-Threideh J., Lehnert L.;  
RT "The sequence of C. elegans cosmid H23L24.";  
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Bristol N2;  
RA Waterston R.;  
RL Submitted (DEC-2002) to the EMBL/GenBank/DBJ databases.  
RN [4]  
RP SEQUENCE FROM N.A.  
RA Gagnon S.N., Hardy I., Desnoyers S.;  
RT "Molecular characterization of poly ADP-ribose glycohydrolases in  
Caenorhabditis elegans.";  
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AC006662; AAF39896.2; -;  
DR EMBL; AF548468; AAN40599.1; -;  
DR WormPep; H23L24.5; CE32685.  
DR InterPro; IPR007724; PARG.  
DR Pfam; PF05028; PARG; 1.  
KW Hypothetical protein.  
SQ SEQUENCE 485 AA; 55574 MW; 86367DDBAB15081F CRC64;  
Query Match 8.9%; Score 458; DB 5; Length 485;  
Best Local Similarity 29.3%; Pred. No. 1.5e-25;  
Matches 142; Conservative 84; Mismatches 160; Indels 98; Gaps 18;







Search completed: May 26, 2004, 18:46:06  
Job time : 81.2527 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 26, 2004, 18:18:25 ; Search time 77.8912 Seconds

(without alignments)  
3953.539 Million cell updates/sec

Title: US-09-302-812-4

Perfect score: 5190

Sequence: 1 MNAGGCEPCTKATRWGAAT.....YHAVSCAETADHSQRTGT 976

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: sp\_archaea.\*

2: sp\_bacteria.\*

3: sp\_fungi.\*

4: sp\_human.\*

5: sp\_invertebrate.\*

6: sp\_mammal.\*

7: sp\_mhc.\*

8: sp\_organelle.\*

9: sp\_phase.\*

10: sp\_plant.\*

11: sp\_rodent.\*

12: sp\_virus.\*

13: sp\_vertebrate.\*

14: sp\_unclassified.\*

15: sp\_rvirus.\*

16: sp\_bacteriaph.\*

17: sp\_archaeap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	5190	100.0	976	4 Q9Y4W7	Q9Y4W7 homo sapien
2	5113	98.5	976	4 Q86W56	Q86W56 homo sapien
3	5108	98.4	976	4 Q7Z742	Q7Z742 homo sapien
4	4635.5	89.3	977	6 Q02776	Q02776 bos taurus
5	4395	84.7	968	11 Q86222	Q86222 mus musculus
6	4381.5	84.4	961	11 Q80YQ6	Q80YQ6 mus musculus
7	4336	83.5	972	11 Q9QIM2	Q9QIM2 rattus norv
8	4165.5	80.3	920	11 Q8CB72	Q8CB72 mus musculus
9	1078	20.8	768	5 Q46043	Q46043 drosophila
10	1073	20.7	723	5 Q960N8	Q960N8 drosophila
11	579	11.2	548	10 Q9SKB3	Q9SKB3 arabidopsis
12	539	10.4	522	10 Q8VIAL	Q8VIAL arabidopsis
13	480.5	9.3	781	5 Q867X0	Q867X0 caenorhabdi
14	473	9.1	764	5 Q19637	Q19637 caenorhabdi
15	453	8.7	485	5 Q9N5L4	Q9N5L4 caenorhabdi
16	340	6.6	364	10 Q9SKB4	Q9SKB4 arabidopsis

17	333.5	6.4	368	5 Q86GI4	Q86GI4 toxoplasma
18	188	3.6	508	12 Q9E234	Q9E234 helicoverpa
19	184	3.5	510	12 Q99GU9	Q99GU9 helicoverpa
20	178	3.4	2081	10 Q3LH98	Q3LH98 arabidopsis
21	175	3.4	1078	5 Q963T1	Q963T1 plasmodium
22	172	3.3	2241	5 Q8IK37	Q8IK37 plasmodium
23	171	3.3	744	5 Q86S96	Q86S96 babesia rod
24	170.5	3.3	4969	11 Q8CF91	Q8CF91 mus musculu
25	170.5	3.3	5165	11 Q8CF92	Q8CF92 mus musculu
26	170	3.3	5322	5 Q9VPL9	Q9VPL9 drosophila
27	169.5	3.3	1274	16 Q8NW31	Q8NW31 staphylococ
28	169	3.3	498	5 Q8MTN8	Q8MTN8 trichinella
29	168	3.2	3167	5 Q17464	Q17464 caenorhabdi
30	167.5	3.2	968	5 Q8I5W9	Q8I5W9 plasmodium
31	167	3.2	1988	5 Q9VU16	Q9VU16 drosophila
32	166.5	3.2	1163	4 Q9UHB7	Q9UHB7 homo sapien
33	166.5	3.2	1472	5 Q8I2A4	Q8I2A4 plasmodium
34	165.5	3.2	571	5 Q8MTN7	Q8MTN7 trichinella
35	165.5	3.2	1274	16 Q99TC3	Q99TC3 staphylococ
36	165.5	3.2	1444	5 Q9VTN2	Q9VTN2 drosophila
37	165.5	3.2	1514	5 Q8SY55	Q8SY55 drosophila
38	165	3.2	838	10 Q9CAK9	Q9CAK9 arabidopsis
39	165	3.2	884	10 Q9CAB4	Q9CAB4 arabidopsis
40	164	3.2	4717	3 Q94248	Q94248 schizosacch
41	163.5	3.2	864	16 Q7ZAM9	Q7ZAM9 leptospira
42	163.5	3.2	1182	5 Q8ID30	Q8ID30 plasmodium
43	163	3.1	349	6 Q8HY98	Q8HY98 centurio se
44	162	3.1	455	5 Q86A18	Q86A18 dictyosteli
45	162	3.1	585	5 Q8IG74	Q8IG74 caenorhabdi

## ALIGNMENTS

### RESULT 1

Q9Y4W7 PRELIMINARY; PRT; 976 AA.  
AC Q9Y4W7  
DT 01-NOV-1999 (TREMELrel. 12, Created)  
DT 01-NOV-1999 (TREMELrel. 12, Last sequence update)  
DT 01-JUN-2003 (TREMELrel. 24, Last annotation update)  
DE POLY(ADP-ribose) glycohydrolase.  
GN HPARG.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP MEDLINE=9380098; PubMed=10449915;  
RX Medline=9380098; AAB61614.1; Jacobson M.K.;  
RA Ane J.C., Apion F., Jacobson E.L., Jacobson M.K.;  
RT "Assignment of the poly (ADP-ribose) glycohydrolase gene (PARG) to  
human chromosome 10q11.23 and mouse chromosome 14B by in situ  
hybridization.";  
RL Cytogenet. Cell Genet. 85:269-270(1999).  
DR EMBL; AF005043; AAB61614.1; -;  
DR Genbank; HGNC:8605; PARG.  
DR GO; GO:0005737; C:cytoplasm; TAS.  
DR GO; GO:0004649; F:poly(ADP-ribose) glycohydrolase activity; TAS.  
DR InterPro; IPR007724; PARG.  
DR Pfam; PF05028; PARG; 1.  
DR KEGG; K01103; PARG; 1.  
SQ SEQUENCE 976 AA; 110903 MW; 42A0F51466FCE7DC CRC64;

Query Match 100.0%; Score 5190; DB 4; Length 976;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 976; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MNAGGCEPCTKATRWGAATTPAASDARSPQRQRLVDPKDAHVFRVPPSPACVPG 60  
Db 1 MNAGGCEPCTKATRWGAATTPAASDARSPQRQRLVDPKDAHVFRVPPSPACVPG 60  
Qy 61 QAGQHRGSATSLVFKQTKITTSWMDTKIGIKTAESLDSKENNNTRIESMMSSVQKDNFYQ 120

Db 61 QAGHRSATSLVFKQKTTISWMDTKGKTAESLSDSKENNTRIESMSSVQKDNFYQ 120  
Qy 121 HNVEKLVNVSQSLDKSLTEKSTQVNLNQHQAAMCKWQNEKGKTEQLLESEPTVTLVPE 180  
Db 121 HNVEKLVNVSQSLDKSLTEKSTQVNLNQHQAAMCKWQNEKGKTEQLLESEPTVTLVPE 180  
Qy 181 QFSNANIDRSQNDHSDTDSEENRDNQOFLTTVKLANAKOTTEDHAREAKSHQKCSKS 240  
Db 181 QFSNANIDRSQNDHSDTDSEENRDNQOFLTTVKLANAKOTTEDHAREAKSHQKCSKS 240  
Qy 241 CHPGEDCASCQDEIDVVPKSPLSVGSSEDTGTSKNDKLIQBSCLGNSPPPEKESEP 300  
Db 241 CHPGEDCASCQDEIDVVPKSPLSVGSSEDTGTSKNDKLIQBSCLGNSPPPEKESEP 300  
Qy 301 ESPMDVDNSKNSCQSEADETSPGFDEQDSSSQTANKPSRFOARDADIEFRKRYSTK 360  
Db 301 ESPMDVDNSKNSCQSEADETSPGFDEQDSSSQTANKPSRFOARDADIEFRKRYSTK 360  
Qy 361 GGEVLHFOFEGESRTGMNDLNAKLPGNISLNVCECRNSKHGKDKSKI TDHLMRLPKA 420  
Db 361 GGEVLHFOFEGESRTGMNDLNAKLPGNISLNVCECRNSKHGKDKSKI TDHLMRLPKA 420  
Qy 421 EDRRKEQWETKHQTERKIPKVPVPHLSPDKKWLGTPIEEMRMRPCGIRLPLRPSANH 480  
Db 421 EDRRKEQWETKHQTERKIPKVPVPHLSPDKKWLGTPIEEMRMRPCGIRLPLRPSANH 480  
Qy 481 TVTIRVDLLRAGEVPKPPFTTHYKDLWKNKHVQMPCEQNLYPVEDENGERTAGSRWELIQ 540  
Db 481 TVTIRVDLLRAGEVPKPPFTTHYKDLWKNKHVQMPCEQNLYPVEDENGERTAGSRWELIQ 540  
Qy 541 TALLNKFTRPQNLKDAILKYNVAYSKKWDFTALIDFWDKVLEAEQAHLVYQSILPDMVKI 600  
Db 541 TALLNKFTRPQNLKDAILKYNVAYSKKWDFTALIDFWDKVLEAEQAHLVYQSILPDMVKI 600  
Qy 601 ALCLPNICTQPIPLLKQKNHSTMSQEQIASILANAFCTTFRNNAKMSSEYSSYPDIN 660  
Db 601 ALCLPNICTQPIPLLKQKNHSTMSQEQIASILANAFCTTFRNNAKMSSEYSSYPDIN 660  
Qy 661 FNRLFEGRSSRKPKEKLTLCYFRRVTEKPTGLVTFTRQSLDFFPEWERCCKPLTRLHV 720  
Db 661 FNRLFEGRSSRKPKEKLTLCYFRRVTEKPTGLVTFTRQSLDFFPEWERCCKPLTRLHV 720  
Qy 721 TYEGTTEENGQGMLOVDFAFRVGGVTSAGLVQBEIRFLINPELLIISRLFTVLDHNEC 780  
Db 721 TYEGTTEENGQGMLOVDFAFRVGGVTSAGLVQBEIRFLINPELLIISRLFTVLDHNEC 780  
Qy 781 LIITGTEQYSEYTGAYETRWRSRSHEDGSDRDCERRCTEIVADALHFRRYLDQFVPEK 840  
Db 781 LIITGTEQYSEYTGAYETRWRSRSHEDGSDRDCERRCTEIVADALHFRRYLDQFVPEK 840  
Qy 841 MRRLNKAYCGFLRPGVSSENLSAVATGNWCGAFGGDARLKALIQILAAAAAERDVVYF 900  
Db 841 MRRLNKAYCGFLRPGVSSENLSAVATGNWCGAFGGDARLKALIQILAAAAAERDVVYF 900  
Qy 901 TFGDSELMRDYIYSHWIFLTERKLTVDGVYKLLRYNEECRNCSTPGPDIKLYFFIYHAV 960  
Db 901 TFGDSELMRDYIYSHWIFLTERKLTVDGVYKLLRYNEECRNCSTPGPDIKLYFFIYHAV 960  
Qy 961 ESCAETADHSGQRTGT 976  
Db 961 ESCAETADHSGQRTGT 976

## RESULT 2

Q86W56 PRELIMINARY; PRT; 976 AA.  
AC Q86W56;  
DT 01-JUN-2003 (TrEMBLrel. 24, Created)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Similar to poly (ADP-ribose) glycohydrolase.  
OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Skin;  
RA Strauberg R.;  
RL Submitted (APR-2003) to the EMBL/GenBank/DBJ databases.  
DR GO:0016787; F:hydrolase activity; IEA.  
DR InterPro: IPR007724; PARG.  
DR Pfam: PF05028; PARG; 1.  
KW Hydrolase.  
SQ SEQUENCE 976 AA; 111109 MW; D6646353C6D0180E CRC64;  
Query Match 98.5%; Score 5113; DB 4; Length 976;  
Best Local Similarity 98.7%; Pred. No. 0;  
Matches 963; Conservative 3; Mismatches 10; Indels 0; Gaps 0;  
Qy 1 MNAGGCEPCTKATRWGAATTSPAASDARSPPSRQRRVLDPKDAHVQVRVPPSPACVPG 60  
Db 1 MNAGGCEPCTKATRWGAATTSPAASDARSPPSRQRRVLDPKDAHVQVRVPPSPACVPG 60  
Qy 61 QAGHRSATSLVFKQKTTISWMDTKGKTAESLSDSKENNTRIESMSSVQKDNFYQ 120  
Db 61 QAGHRSATSLVFKQKTTISWMDTKGKTAESLSDSKENNTRIESMSSVQKDNFYQ 120  
Qy 121 HNVEKLVNVSQSLDKSLTEKSTQVNLNQHQAAMCKWQNEKGKTEQLLESEPTVTLVPE 180  
Db 121 HNVEKLVNVSQSLDKSLTEKSTQVNLNQHQAAMCKWQNEKGKTEQLLESEPTVTLVPE 180  
Qy 181 QFSNANIDRSQNDHSDTDSEENRDNQOFLTTVKLANAKOTTEDHAREAKSHQKCSKS 240  
Db 181 QFSNANIDRSQNDHSDTDSEENRDNQOFLTTVKLANAKOTTEDHAREAKSHQKCSKS 240  
Qy 241 CHPGEDCASCQDEIDVVPKSPLSVGSSEDTGTSKNDKLIQBSCLGNSPPPEKESEP 300  
Db 241 CHPGEDCASCQDEIDVVPKSPLSVGSSEDTGTSKNDKLIQBSCLGNSPPPEKESEP 300  
Qy 301 ESPMDVDNSKNSCQSEADETSPGFDEQDSSSQTANKPSRFOARDADIEFRKRYSTK 360  
Db 301 ESPMDVDNSKNSCQSEADETSPGFDEQDSSSQTANKPSRFOARDADIEFRKRYSTK 360  
Qy 361 GGEVLHFOFEGESRTGMNDLNAKLPGNISLNVCECRNSKHGKDKSKI TDHLMRLPKA 420  
Db 361 GGEVLHFOFEGESRTGMNDLNAKLPGNISLNVCECRNSKHGKDKSKI TDHLMRLPKA 420  
Qy 421 EDRRKEQWETKHQTERKIPKVPVPHLSPDKKWLGTPIEEMRMRPCGIRLPLRPSANH 480  
Db 421 EDRRKEQWETKHQTERKIPKVPVPHLSPDKKWLGTPIEEMRMRPCGIRLPLRPSANH 480  
Qy 481 TVTIRVDLLRAGEVPKPPFTTHYKDLWKNKHVQMPCEQNLYPVEDENGERTAGSRWELIQ 540  
Db 481 TVTIRVDLLRAGEVPKPPFTTHYKDLWKNKHVQMPCEQNLYPVEDENGERTAGSRWELIQ 540  
Qy 541 TALLNKFTRPQNLKDAILKYNVAYSKKWDFTALIDFWDKVLEAEQAHLVYQSILPDMVKI 600  
Db 541 TALLNKFTRPQNLKDAILKYNVAYSKKWDFTALIDFWDKVLEAEQAHLVYQSILPDMVKI 600  
Qy 601 ALCLPNICTQPIPLLKQKNHSTMSQEQIASILANAFCTTFRNNAKMSSEYSSYPDIN 660  
Db 601 ALCLPNICTQPIPLLKQKNHSTMSQEQIASILANAFCTTFRNNAKMSSEYSSYPDIN 660  
Qy 661 FNRLFEGRSSRKPKEKLTLCYFRRVTEKPTGLVTFTRQSLDFFPEWERCCKPLTRLHV 720  
Db 661 FNRLFEGRSSRKPKEKLTLCYFRRVTEKPTGLVTFTRQSLDFFPEWERCCKPLTRLHV 720  
Qy 721 TYEGTTEENGQGMLOVDFAFRVGGVTSAGLVQBEIRFLINPELLIISRLFTVLDHNEC 780  
Db 721 TYEGTTEENGQGMLOVDFAFRVGGVTSAGLVQBEIRFLINPELLIISRLFTVLDHNEC 780  
Qy 781 LIITGTEQYSEYTGAYETRWRSRSHEDGSDRDCERRCTEIVADALHFRRYLDQFVPEK 840  
Db 781 LIITGTEQYSEYTGAYETRWRSRSHEDGSDRDCERRCTEIVADALHFRRYLDQFVPEK 840

```

Db      781 LIITGTEQSEYTGAEYVNSRSHEDGSDRDPQRRCTEIVADLHFRYLDQFVPEK 840
Qy      841 MRRELKAYCGFLRPGVSSNLSAVATGNWCGAFGGDARLKALIQILAAAAAERDVVYF 900
Db      841 MRRELKAYCGFLRPGVSSNLSAVATGNWCGAFGGDARLKALIQILAAAAAERDVVYF 900
Qy      901 TFGDSELMRDYISNMHIFLTERKLTGVDVYKLLRYNEECNCSSTGPDIKLPPFIYHAV 960
Db      901 TFGDSELMRDYISNMHIFLTERKLTGVDVYKLLRYNEECNCSSTGPDIKLPPFIYHAV 960
Qy      961 ESCAETADHSGQRTGT 976
Db      961 ESCAETADHSGQRTGT 976

RESULT 3
Q72742 ID Q72742 PRELIMINARY; PRT; 976 AA.
AC Q72742;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Testis;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.J., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heish F.,
RA Dratchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Stalska U., Smallos D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RA "Generation and initial analysis of more than 15,000 full-length human
RA and mouse cDNA sequences."
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Testis;
RA Strausberg R.;
RL Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC052966; AAHS2966.1; -.
KW Hypothetical protein.
SQ SEQUENCE 976 AA; 111090 MW; 47C8EE1826C4A74A CRC64;

```

```

Query Match 98.4%; Score 5108; DB 4; Length 976;
Best Local Similarity 98.6%; Pred. No. 0;
Matches 962; Conservative 3; Mismatches 11; Indels 0; Gaps 0;

```

```

Qy      1 MNAGGCEPCTKATGWAATTSPPASDARSPPSRORRVLDPKDAHQVFPVPPSPACVPG 60
Db      1 MNAGGCEPCTKATGWAATTSPPASDARSPPSRORRVLDPKDAHQVFPVPPSPACVPG 60
Qy      61 QAGQHRGATSILVFQKTIITSMMDTKGKTAESLDSKENNTRIESMSSVQKDNFYQ 120
Db      61 RAGQHRGATSILVFQKTIITSMMDTKGKTAESLDSKENNTRIESMSSVQKDNFYQ 120
Qy      121 HNVEKLVNSQLSDKSLTEKSTQYLNQHTAAMCKWQNEGKHTEQLLESEPTVTLVPE 180

```

```

Db      121 HNVEKLVNSQLSDKSLTEKSTQYLNQHTAAMCKWQNEGKHTEQLLESEPTVTLVPE 180
Qy      181 QFSNANIDRSQNDHSDTDSENRDNOQFLTTVKLANAKQTTDEHARAKSHQKSKS 240
Db      181 QFSNANIDRSQNDHSDTDSENRDNOQFLTTVKLANAKQTTDEHARAKSHQKSKS 240
Qy      241 CHGGEPCASQDQDEIDVVPKSPLSVDGSEVGTGSKNDKLIROESCLGNSPPEKESEP 300
Db      241 CDGGEPCASQDQDEIDVVPKSPLSVDGSEVGTGSKNDKLIROESCLGNSPPEKESEP 300
Qy      301 ESPMDVDNKNKNSQDSEADEETSPGDEQEDGSSSTANKPSRPFQARDADIEFRKRYSTK 360
Db      301 ESPMDVDNKNKNSQDSEADEETSPGDEQEDGSSSTANKPSRPFQARDADIEFRKRYSTK 360
Qy      361 GGEVRLHFQFEGGESRTGMNDLNAKLPGNISLNVCRNSKQHGKDKSKITDHLMLRPLKA 420
Db      361 GGEVRLHFQFEGGESRTGMNDLNAKLPGNISLNVCRNSKQHGKDKSKITDHLMLRPLKA 420
Qy      421 EDRRKEQWETKHQRTERKIPKYPPHLSPPDKKWLGTPIEMRRMPRCGIELPLRPSANH 480
Db      421 EDRRKEQWETKHQRTERKIPKYPPHLSPPDKKWLGTPIEMRRMPRCGIELPLRPSANH 480
Qy      481 TVTIRVDLLRAGEVPKPPFPHYKDLWKNHVKMPCSEONLYPVEDENGERTAGSRWELIQ 540
Db      481 TVTIRVDLLRAGEVPKPPFPHYKDLWKNHVKMPCSEONLYPVEDENGERTAGSRWELIQ 540
Qy      541 TALLNKFTRPQNLKDAIKYNVAYSKKWDFTALIDFWDKVLBAEAQHLVQSILPDMVKI 600
Db      541 TALLNKFTRPQNLKDAIKYNVAYSKKWDFTALIDFWDKVLBAEAQHLVQSILPDMVKI 600
Qy      601 ALCLPNICTQPIPLLKQKNHSITMSQEQIATSLANAFCTFPPRNAMKSEYSSYPDIIN 660
Db      601 ALCLPNICTQPIPLLKQKNHSITMSQEQIATSLANAFCTFPPRNAMKSEYSSYPDIIN 660
Qy      661 FNRLFEGRSRKKPEKLTFLCYFRRVTEKPTGLVTFTRQSLDFPPEWCEKEPLTRLHV 720
Db      661 FNRLFEGRSRKKPEKLTFLCYFRRVTEKPTGLVTFTRQSLDFPPEWCEKEPLTRLHV 720
Qy      721 TYEGTIEENGQGLQVDFANRFVGGVTSAGLVQVEEIRFLINPELLIISRLTFTEVDHNEC 780
Db      721 TYEGTIEENGQGLQVDFANRFVGGVTSAGLVQVEEIRFLINPELLIISRLTFTEVDHNEC 780
Qy      781 LIITGTEQSEYTGAEYVNSRSHEDGSDRDPQRRCTEIVADLHFRYLDQFVPEK 840
Db      781 LIITGTEQSEYTGAEYVNSRSHEDGSDRDPQRRCTEIVADLHFRYLDQFVPEK 840
Qy      841 MRRELKAYCGFLRPGVSSNLSAVATGNWCGAFGGDARLKALIQILAAAAAERDVVYF 900
Db      841 MRRELKAYCGFLRPGVSSNLSAVATGNWCGAFGGDARLKALIQILAAAAAERDVVYF 900
Qy      901 TFGDSELMRDYISNMHIFLTERKLTGVDVYKLLRYNEECNCSSTGPDIKLPPFIYHAV 960
Db      901 TFGDSELMRDYISNMHIFLTERKLTGVDVYKLLRYNEECNCSSTGPDIKLPPFIYHAV 960
Qy      961 ESCAETADHSGQRTGT 976
Db      961 ESCAETADHSGQRTGT 976

RESULT 4
Q02776 ID Q02776 PRELIMINARY; PRT; 977 AA.
AC Q02776;
DT 01-JUL-1997 (TrEMBLrel. 04, Created)
DT 01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Poly(ADP-ribose) glycohydrolase.
GN BPARG.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bos.

```

OX NCBI\_TaxID=9913;  
 RN SEQUENCE FROM N.A.  
 RP MEDLINE=97277328; PubMed=9115250;  
 RX Lin W., Ame J.C., Abouli-Ela N., Jacobson E.L., Jacobson M.K.;  
 RA "Isolation and characterization of the cDNA encoding bovine poly (ADP-  
 RT ribose) glycohydrolase."  
 RL J. Biol. Chem. 272:11895-11901 (1997).  
 RR EMBL; U78975; AAB53370.1; -;  
 DR GO; GO:0016787; F:hydrolase activity; IEA.  
 DR InterPro; IPR007724; PARG.  
 DR Pfam; PF05028; PARG; 1.  
 KW Hydrolase.  
 SQ SEQUENCE 977 AA; 110837 MW; 87D2100F979DF377 CRC64;  
  
 Query Match 89.3%; Score 4635.5; DB 6; Length 977;  
 Best Local Similarity 89.5%; Pred. No. 0;  
 Matches 874; Conservative 32; Mismatches 69; Indels 1; Gaps 1;  
  
 QY 1 MNAGGCEPCTKATRWGAATTS-PAASDARFPQRORRVLPDKAHVQFRVPPSPACVP 59  
 DB 1 MSAGFGCEPCTKRPRWDAATSPPAASDARFPQRORRVLPDKAHVQFRVPPSPACVP 60  
  
 QY 60 GOAGHGRGSAATSLVFKQKTIISWMDTKGKTAESLDSKNNNTRIESMSSVQKDNFY 119  
 DB 61 GRAGHGRGSAATSLVFKQKTIISWMDTKGKTAESLDSKNNNTRIESMSSVQKDNFY 120  
  
 QY 120 QHNYEKLNVVQSLSLDSKSLTEKSTQYLNQHOHTAAKCKWQNEKTEQLLESEPTVTLVP 179  
 DB 121 QHNYEKLNVVQSLSLDSKSLTEKSTQYLNQHOHTAAKCKWQNEKTEQLLESEPTVTLVP 180  
  
 QY 180 EQFSNANIDRSPQNDHSDTSEENRDNQOFLTTVKLANAKQTTEDEHARAKSHQKCSK 239  
 DB 181 EQFSNANIDRSPQNDHSDTSEENRDNQOFLTTVKLANAKQTTEDEHARAKSHQKCSK 240  
  
 QY 240 SCHGEDCASCOQDIDVVPKSPISDVGSSEDTGSGKNDKLIQESCLGNSPPPEKSE 299  
 DB 241 ACHPAEACAGCQOETDVVVSPLSDTGSDDTGVGKNNANRQESLGSNPPPEKSE 300  
  
 QY 300 PESPMVDVNSKNSCODSEADETSFGFDEQEDGSSQATANKPSRFQARDADIEFRKRYST 359  
 DB 301 PESPMVDVNSKNSCODSEADETSFGFDEQEDGSSQATANKPSRFQARDADIEFRKRYST 360  
  
 QY 360 KGGEVRLHFQEGGESRTGMNDLNAKLPGNTSSLNVECRNSKHQKSKITDHLMLPK 419  
 DB 361 KGGEVRLHFQEGGESRTGMNDLNAKLPGNTSSLNVECRNSKHQKSKITDHLMLPK 420  
  
 QY 420 AEDRKEQWETKHQTERKIPKYPVPHLSPDKKWLGTPIEEMRMPCGIRLPLRPSAN 479  
 DB 421 AEDRKEQWETKHQTERKIPKYPVPHLSPDKKWLGTPIEEMRMPCGIRLPLRPSAN 480  
  
 QY 480 HTVTIRVDLLRAGEVPKPPPTHYKOLWKNKVKMPCSEONLYPVDENGERTAGSRWELI 539  
 DB 481 HTVTIRVDLLRAGEVPKPPPTHYKOLWKNKVKMPCSEONLYPVDENGERTAGSRWELI 540  
  
 QY 540 QTALLNKTRPQNLDAILKYNVAYSKWDFALIDFWDKVLEEAQHLVOSIILPDMVK 599  
 DB 541 QTALLNKTRPQNLDAILKYNVAYSKWDFALIDFWDKVLEEAQHLVOSIILPDMVK 600  
  
 QY 600 IALCLPNICTQPIPLKQKNSHITMSQIASLANAFCTFPRNNAKMSSEYSYDPI 659  
 DB 601 IALCLPNICTQPIPLKQKNSHITMSQIASLANAFCTFPRNNAKMSSEYSYDPI 660  
  
 QY 660 NFNRLEFEGSSRRKPEKLTLCYFRVATEKXPTGLVTFTRQSLDPPPEWERCEKPLTRLH 719  
 DB 661 NFNRLEFEGSSRRKPEKLTLCYFRVATEKXPTGLVTFTRQSLDPPPEWERCEKPLTRLH 720  
  
 QY 720 VTYEGTIENGQGLQVDPAFNVGGVTSAGLVQBEIRFLINPELLISRLTEVLHDNE 779  
 DB 721 VTYEGTIENGQGLQVDPAFNVGGVTSAGLVQBEIRFLINPELLISRLTEVLHDNE 780  
  
 QY 780 CLIIITCTEYSEVTGYAETVYRWARSHERDDWQRRTEIIVDAIHFRRLYDQFVPE 839

DB 781 CLIIITCTEYSEVTGYAETVYRWARSHERDDWQRRTEIIVDAIHFRRLYDQFVPE 840  
 QY 840 XMRRELKAYCGFLRFGVSSENISAVATGNWGCAGGADARLKALIQILAAAAAERDVVY 899  
 DB 841 XMRRELKAYCGFLRFGVSSENISAVATGNWGCAGGADARLKALIQILAAAAAERDVVY 900  
 QY 900 FTGDSSELMDIYSMEIFLTERKLTVDGVYKLLRYNNEECRNCSTPGPDIKLYPFYIHA 959  
 DB 901 FTGDSSELMDIYSMEIFLTERKLTVDGVYKLLRYNNEECRNCSTPGPDIKLYPFYIHA 960  
 QY 960 VESCAETADHSQRTG 975  
 DB 961 VESCTOTTNPQRTG 976  
  
 RESULT 5  
 O88622 PRELIMINARY; PRT; 968 AA.  
 AC O88622;  
 DT 01-NOV-1998 (TrEMBLrel. 08, Created)  
 DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
 DE Poly (ADP-ribose) glycohydrolase.  
 GN PARG.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OC NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=99380098; PubMed=10449915;  
 RA Ame J.C., Aplou F., Jacobson E.L., Jacobson M.K.;  
 RT "Assignment of the poly (ADP-ribose) glycohydrolase gene (PARG) to  
 RT human chromosome 10q11.23 and mouse chromosome 14B by in situ  
 RT hybridization."  
 RL Cytogenet. Cell Genet. 85:269-270 (1999).  
 DR EMBL; AF079557; AAC28735.1; -;  
 DR MGD; MGI:1347094; Parg.  
 DR GO; GO:0016787; F:hydrolase activity; IEA.  
 DR InterPro; IPR007724; PARG.  
 DR Pfam; PF05028; PARG; 1.  
 KW Hydrolase.  
 SQ SEQUENCE 968 AA; 109169 MW; 3981DA4809E1E007 CRC64;  
  
 Query Match 84.7%; Score 4395; DB 11; Length 968;  
 Best Local Similarity 85.7%; Pred. No. 1.1e-306;  
 Matches 837; Conservative 52; Mismatches 78; Indels 10; Gaps 6;  
  
 QY 1 MNAGGCEPCTKATRWGAATTS-PAASDARFPQRORRVLPDKAHVQFRVPPSPACVP 59  
 DB 1 MSAGFGCEPCTKA-RWGAAGTSAPTASDSRSPQRORRVLPDKAHVQFRVPPSPACVS 59  
  
 QY 60 GOAGHGRGSAATSLVFKQKTIISWMDTKGKTAESLDSKNNNTRIESMSSVQKDNFY 119  
 DB 60 GRAGHGRGSAATSLVFKQKTIISWMDTKGKTAESLDSKNNNTRIESMSSVQKDNFY 116  
  
 QY 120 QHNYEKLNVVQSLSLDSKSLTEKSTQYLNQHOHTAAKCKWQNEKTEQLLESEPTVTLVP 179  
 DB 117 PHKVEKLENVQNLNDKSPTEKSSQYLNQQTASVCKWQNEKHAQLLASEPPAGTLP 176  
  
 QY 180 EQFSNANIDRSPQNDHSDTSEENRDNQOFLTTVKLANAKQTTEDEHARAKSHQKCSK 239  
 DB 177 KQLSNANIGSPHTDDHSDTDEEDRDNQOFLTPKLANTKPTVGDQAR---SNCKCSG 233  
  
 QY 240 SCHGEDCASCOQDIDVVPKSPISDVGSSEDTGSGKNDKLIQESCLGNSPPPEKSE 299  
 DB 234 SRQSVKDCGTCQOQEVVLPESPLSDVGAEDIGTPKNDKLTGQESSIGSSPPPEKSE 293  
  
 QY 300 PESPMVDVNSKNSCODSEADETSFGFDEQEDGSSQATANKPSRFQARDADIEFRKRYST 359  
 DB 294 PESPMVDVNSKNSCODSEADETSFGFDEQEDGSSQATANKPSRFQARDADIEFRKRYST 352  
  
 QY 360 KGGEVRLHFQEGGESRTGMNDLNAKLPGNTSSLNVECRNSKHQKSKITDHLMLPK 419

```

Db      353  KGSEVRLHFQFE--GENNAGTSDLNKPSGNSSLNVECRSSKHGKRDSDKITDHFMRISK 411
QY      420  AEDRRKEQWETKHQRTERKIPKYVPHLSPDKKWLGTPIEEMRMPCRGIRLPLLRPSAN 479
Db      412  SEDRRKEQCEVHRQTERKIPKYIPNLPPEKKWLGTPIEEMRMPCRGIRLPLLRPSAS 471
QY      480  HTVTIRVDLLRAGEVPEKPTTHYKDIWDNKHVMPQSEONLYPVEDENGERTAGSRWELI 539
Db      472  HTVTIRVDLLRAGEVPEKPTTHYKDIWDNKHVMPQSEONLYPVEDENGERTAGSRWELI 531
QY      540  QTALLNKFTRPQNLKDALIKYNVAYSKKWDFTALIDFWDKVLBEAQAHLIYQSILPDMVK 599
Db      532  QTALLNKFTRPQNLKDALIKYNVAYSKKWDFTALIDFWDKVLBEAQAHLIYQSILPDMVK 591
QY      600  IALCLPNICTQPIPLLKQKNHSHITMSQEOIASLLANAFCTPPRRNAKMSYSSYPDI 659
Db      592  IALCLPNICTQPIPLLKQKNHSHITMSQEOIASLLANAFCTPPRRNAKMSYSSYPDI 651
QY      660  NFNRLEFEGSSRRKPEKLTILFCYFRVRVTEKKPTGLVTFTRQSLDPPWERCCEKPLTRLH 719
Db      652  NFNRLEFEGSSRRKPEKLTILFCYFRVRVTEKKPTGLVTFTRQSLDPPWERCCEKPLTRLH 711
QY      720  VTVEGTIEENGQMLQVDPANRFGGGVTSAGLVQBEIRFLINPELIIISRLFTVLDHNE 779
Db      712  VTVEGTIEENGQMLQVDPANRFGGGVTSAGLVQBEIRFLINPELIIISRLFTVLDHNE 771
QY      780  CLIIITGEQYSEYTGVAETVRWSRSHEDGSDRDCERRCTEIVAIIDALHFRYLDQFVPE 839
Db      772  CLIIITGEQYSEYTGVAETVRWSRSHEDGSDRDCERRCTEIVAIIDALHFRYLDQFVPE 831
QY      840  KMRELNKAYCGFLRPGVPSSENLSAVATGNWGCAGFGGDARLKALIQILAAAAAERDVVY 899
Db      832  KMRELNKAYCGFLRPGVPSSENLSAVATGNWGCAGFGGDARLKALIQILAAAAAERDVVY 891
QY      900  FTGDSSELMRDIYSMHIFLTERKLTVDGVYKLLIRYNEECRCNSTPGPDIKLYPFIYHA 959
Db      892  FTGDSSELMRDIYSMHIFLTERKLTVDGVYKLLIRYNEECRCNSTPGPDIKLYPFIYHA 951
QY      960  VESCAETADHSQORTGT 976
Db      952  VESSAETDMPGQKAGT 968

```

## RESULT 6

```

Q80YQ6  Q80YQ6  PRELIMINARY;  PRT;  961 AA.
ID      AC      Q80YQ6
DT      01-JUN-2003 (TrEMBLrel. 24, Created)
DT      01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT      01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE      Poly (ADP-ribose) glycohydrolase.
OS      Mus musculus (Mouse);
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
ON      NCBI_TaxID=10090;
RX      [1]
RP      SEQUENCE FROM N. A.
RC      STRAIN=C57BL/6; TISSUE=Brain;
RA      Strauberg R.;
RL      Submitted (APR-2003) to the EMBL/GenBank/DBJ databases.
DR      EMBL; BC050892; AAHS0892.1; -
DR      GO; GO:0016787; F:hydrolase activity; IEA.
DR      InterPro; IPR007724; PARG.
DR      Pfam; PF05028; PARG; 1.
KW      Hydrolase.
SQ      SEQUENCE 961 AA; 108580 MW; 1558F120C367CF63 CRC64;

```

Query Match 84.4%; Score 4381.5; DB 11; Length 961;  
 Best Local Similarity 86.0%; Pred. No. 9.7e-306;  
 Matches 833; Conservative 50; Mismatches 77; Indels 9; Gaps 5;

QY 1 MNAGGCEPCTKATRWGAA--TTPAASDARSFQRQRRVLDPKDAHVQFRVPPSPACPV 59

```

Db      1  MSAGPGWEPCTKPRWGAAGTSAPTASRSRSPGQRRRLDPADAPVQFRVPPSPACV 60
QY      60  GQAGQRGSATSLVFKQKTIITSMMDTKGKTAESESLDSKENNTRIESMMSSVQKDNFY 119
Db      61  GRAGPHRGNATSVFKQKTIITSMMDTKGKPTAESE--SKENNTRIDSMSSVQKDNFY 117
QY      120  QNRVRLNVNVSOLDKSLTEKSTQVILNQHOTAAMCKWQNEKGKHTFOLLESPQITVLVP 179
Db      118  PHKVEKLENVPLNDKSPTEKSYQLNQOQTASVCKWQNEGKHAQQLLASEPPAGTPLP 177
QY      180  EQFSNANIDRSQNDHSDTDSEENRDNQOFLTIVKLANAKOTTEDHAREAKSHQKCSK 239
Db      178  KQLSNANIGQSPTDHDSDTDHEEDRDNQOFLTPIKLANTKVTVGQQR--SNCKCSG 234
QY      240  SCHPGECDACSCQDEIDVVPKPSLSDVGSDEVTGSKDNKILRQSCIGNSPPFKESE 299
Db      235  SRQSVKDCCTGCQEEVDVLPESPLSDVGAEDITGPKNDNKLGTQSSSLGDSPPFKESE 294
QY      300  PSPMDVDNKSQCQSEADDETSFGFDEQEDSSSQTANKPSRFOARDADIEFRKRYST 359
Db      295  PSPMDVDNKSQCQSEADDETSFPVFDEQDD--RSSQTANKLSSCCQAREADGOLKRYLT 353
QY      360  KGSEVRLHFQFE--GENNAGTSDLNKPSGNSSLNVECRSSKHGKRDSDKITDHFMRISK 419
Db      354  KGSEVRLHFQFE--GENNAGTSDLNKPSGNSSLNVECRSSKHGKRDSDKITDHFMRISK 412
QY      420  AEDRRKEQWETKHQRTERKIPKYVPHLSPDKKWLGTPIEEMRMPCRGIRLPLLRPSAN 479
Db      413  SEDRRKEQCEVHRQTERKIPKYIPNLPPEKKWLGTPIEEMRMPCRGIRLPLLRPSAS 472
QY      480  HTVTIRVDLLRAGEVPEKPTTHYKDIWDNKHVMPQSEONLYPVEDENGERTAGSRWELI 539
Db      473  HTVTIRVDLLRAGEVPEKPTTHYKDIWDNKHVMPQSEONLYPVEDENGERTAGSRWELI 532
QY      540  QTALLNKFTRPQNLKDALIKYNVAYSKKWDFTALIDFWDKVLBEAQAHLIYQSILPDMVK 599
Db      533  QTALLNKFTRPQNLKDALIKYNVAYSKKWDFTALIDFWDKVLBEAQAHLIYQSILPDMVK 592
QY      600  IALCLPNICTQPIPLLKQKNHSHITMSQEOIASLLANAFCTPPRRNAKMSYSSYPDI 659
Db      593  IALCLPNICTQPIPLLKQKNHSHITMSQEOIASLLANAFCTPPRRNAKMSYSSYPDI 652
QY      660  NFNRLEFEGSSRRKPEKLTILFCYFRVRVTEKKPTGLVTFTRQSLDPPWERCCEKPLTRLH 719
Db      653  NFNRLEFEGSSRRKPEKLTILFCYFRVRVTEKKPTGLVTFTRQSLDPPWERCCEKPLTRLH 712
QY      720  VTVEGTIEENGQMLQVDPANRFGGGVTSAGLVQBEIRFLINPELIIISRLFTVLDHNE 779
Db      713  VTVEGTIEENGQMLQVDPANRFGGGVTSAGLVQBEIRFLINPELIIISRLFTVLDHNE 772
QY      780  CLIIITGEQYSEYTGVAETVRWSRSHEDGSDRDCERRCTEIVAIIDALHFRYLDQFVPE 839
Db      773  CLIIITGEQYSEYTGVAETVRWSRSHEDGSDRDCERRCTEIVAIIDALHFRYLDQFVPE 832
QY      840  KMRELNKAYCGFLRPGVPSSENLSAVATGNWGCAGFGGDARLKALIQILAAAAAERDVVY 899
Db      833  KMRELNKAYCGFLRPGVPSSENLSAVATGNWGCAGFGGDARLKALIQILAAAAAERDVVY 892
QY      900  FTGDSSELMRDIYSMHIFLTERKLTVDGVYKLLIRYNEECRCNSTPGPDIKLYPFIYHA 959
Db      893  FTGDSSELMRDIYSMHIFLTERKLTVDGVYKLLIRYNEECRCNSTPGPDIKLYPFIYHA 952
QY      960  VESCAETAD 968
Db      953  VESSAETTD 961

```

## RESULT 7

Q9QYM2 Q9QYM2 PRELIMINARY; PRT; 972 AA.  
 ID AC Q9QYM2  
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)

DT 01-MAY-2000 (TREMELrel. 13, Last sequence update)  
 DT 01-JUN-2000 (TREMELrel. 24, Last annotation update)  
 DE Poly(ADP-ribose) glycohydrolase.  
 OS Rattus norvegicus (Rat).  
 GN Rattus norvegicus (Rat).  
 OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BUF; TISSUE=Colon;  
 RA Shimokawa T., Masutani M., Nagasawa S., Nozaki T., Ikota N., Araki S.,  
 Aoki Y., Nakagawa H., Sugimura T.;  
 RT "Isolation and cloning of rat poly(ADP-ribose) glycohydrolase  
 (Parg).";  
 RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AB019366; BAA87901.1; -.  
 DR GO; GO:0016787; F:hydrolase activity; IEA.  
 DR InterPro; IPR007724; PARG.  
 DR Pfam; PF05028; PARG; 1.  
 KW Hydrolase.  
 SQ SEQUENCE. 972 AA; 109002 MW; 3DED7885684P59E9 CRC64;

Query Match 83.5%; Score 4336; DB 11; Length 972;  
 Best Local Similarity 84.0%; Pred. No. 1.8e-302;  
 Matches 822; Conservative 56; Mismatches 91; Indels 10; Gaps 5;

QY 1 MNAGPCECTKATRWGAA-TTSPAASDARSPPSRQRRVLDPKDAHVOFRVPPSPACVP 59  
 DB 1 MNAGPCECTKTRPRWGAAGTAPTASDSRSPPGRQKRVLDPKDAPVQVRVPPSSACVS 60  
 QY 60 GAGOHGSGATSLVFKOKTITSMWDTGKTKTAESLSLSDSKENNTRIEISMSSVQKDNFY 119  
 DB 61 GRAGPHRGSVTFVFKQKITTTWMDTKGPKTAESE---SKENNTFTDPMSSVQKDNFY 117  
 QY 120 QHNVEKLVNVSQSLSDKSLTEKSTOYLNOHQTAAMCKWQNEGKHTEQLLESPPQTVLVP 179  
 DB 118 PHKVEKLVNVPQLNDKSPTEKSTPYLNQQTAGVCKWHSAGERAQISASEPSAVTQAP 177  
 QY 180 EGFSNANIDRSQNDHSDTDEEERDNOQFITTVKLANAKOTTEDEHAREAKSHOKCSK 239  
 DB 178 KQLSNANIDQSPPTDGHSDTDEERDNOQFITPVKLANAKOTVGGQAR---SNCKCSA 234  
 QY 240 SCHPGDCCASQODEIDVVPKPSLDVSGSDVGTGSKNDKNTKIRQBSCLGNSPPFEKSE 299  
 DB 235 SCQCGQDCCAGCQCEADVLPESPLSDVGAEDTGTGSKNDKNTKIRQBSCLGNSPPFEKSE 294  
 QY 300 PESPMDVNSKNSQDSSEADEETSPGFDEQ--EDGSSQOTANKPSRFQARDADIEFRKY 357  
 DB 295 PESPMDVNSKTSQDSSEADEEASPVFDEQDDQDDRSSQTANKLSSRQAREVDGDLRKEY 354  
 QY 358 STKGGEVRLHFQEGEGESTGMNDINAKLPGNLSSLNVECSKQHGKDSKITDHLML 417  
 DB 355 LTKGSEIRLHFQEGEG-SNAGTSDLNKAPSGNSLNVDSRSGKQHGKDSKITDHFVKI 413  
 QY 418 PEAEDRRKQEWETKHOKTERKIPKYVPHLSPPDKWLGTPIEEMRMPPRCIGRLPLRFS 477  
 DB 414 PKSEDRRKQCEVHRQAEKIPKYVPPNLPPDKWLGTPIEEMRMPPRCIGRLPLRFS 473  
 QY 478 ANHTVTRVDLLRAGEVPKPPFTHYKDLMDNKHVKMPCSEQNTLYPVEDENGERTAGSRWE 537  
 DB 474 ASHTVTRVDLLRAGEVPKPPFTHYKDLMDNKHVKMPCSEQNTLYPVEDENGERTAGSRWE 533  
 QY 538 LIQTALLNKTRQNLKQDALILKNVAYSKKWDFTALIDFWDKYLEAEAOHLQSLPDM 597  
 DB 534 LIQTALLNKTRQNLKQDALILKNVAYSKKWDFTALVDFWDKYLEAEAOHLQSLPDM 593  
 QY 598 VKTALCLPNICTOPIPLLKQKNHSHITMSQEQIASLIANAFCTFFPRNNAKMSSEYSSYP 657  
 DB 594 VKTALCLPNICTOPIPLLKQKNHSHVITMSQEQIASLIANAFCTFFPRNNAKMSSEYSSYP 653  
 QY 658 DINFNRLFEGRSSRKPEKLTLCYFRRVTEKPTGLVTFTRQSLDEDPPWERCDKPLTR 717

DB 654 DINFNRLFEGRSSRKPEKLTLCYFRRVTEKPTGLVTFTRQSLDEDPPWERCDKPLTR 713  
 QY 718 LHVTYEGTIEENGQCMQVDFANRFVGGVTSAGLVOBEIRFLINPELIIISRLFTFVLDH 777  
 DB 714 LHVTYEGTIEENGQCMQVDFANRFVGGVTSAGLVOBEIRFLINPELIIISRLFTFVLDH 773  
 QY 778 NECLIIITGTEQSYVTGYAETRWRSRSHEDSGERDDCERRCTEIVAIIDALHFRRLVDQPV 837  
 DB 774 NECLIIITGTEQSYVTGYAETRWRSRSHEDSGERDDCERRCTEIVAIIDALHFRRLVDQPV 833  
 QY 838 PEKMRRELKAYCGFLRPGVSSENLSAVATGNWCGAFGGDARLKALIQILAAAAAERDV 897  
 DB 834 PEKVRRELKAYCGFLRPGVPPENLSAVATGNWCGAFGGDARLKALQLAAAAAERDV 893  
 QY 898 VYFTFGDSELMRDYISMHIFLTERKLTGVGDVYKILIRYNEECRNCSTFGPDIKLYPIFY 957  
 DB 894 VYFTFGDSELMRDYISMHIFLTERKLVNGKYVRLRLRYRECRDCSSPGPDTKLYPIFY 953  
 QY 958 HAVESCAETADHSQORTGT 976  
 DB 954 HAAESSAETSQPGORTGT 972

RESULT 8  
 Q8CB72 PRELIMINARY; PRT; 920 AA.  
 AC Q8CB72;  
 DT 01-MAR-2003 (TREMELrel. 23, Created)  
 DT 01-MAR-2003 (TREMELrel. 23, Last sequence update)  
 DT 01-JUN-2003 (TREMELrel. 24, Last annotation update)  
 DE Poly.  
 GN PARG.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Bone;  
 RX MEDLINE=22354683; PubMed=12466851;  
 RA The FANTOM Consortium,  
 RA the RIKEN Genome Exploration Research Group Phase I & II Team;  
 RT "Analysis of the mouse transcriptome based on functional annotation of  
 60,770 full-length cDNAs.";  
 RL Nature 420:563-573 (2002).  
 DR EMBL; AK036656; BAC29519.1; -.  
 DR MGD; MGI:1347094; Parg.  
 DR InterPro; IPR007724; PARG.  
 DR Pfam; PF05028; PARG; 1.  
 SQ SEQUENCE. 920 AA; 103799 MW; 379E779E121B09A1 CRC64;

Query Match 80.3%; Score 4165.5; DB 11; Length 920;  
 Best Local Similarity 85.5%; Pred. No. 3e-290;  
 Matches 793; Conservative 51; Mismatches 75; Indels 9; Gaps 5;

QY 1 MNAGPCECTKATRWGAA-TTSPAASDARSPPSRQRRVLDPKDAHVOFRVPPSPACVP 59  
 DB 1 MSAGPCECTKTRPRWGAAGTAPTASDSRSPPGRQKRVLDPKDAPVQVRVPPSPACVS 60  
 QY 60 GAGOHGSGATSLVFKOKTITSMWDTGKTKTAESLSLSDSKENNTRIEISMSSVQKDNFY 119  
 DB 61 GRAGPHRGSVTFVFKQKITTTWMDTKGPKTAESE---SKENNTFTDPMSSVQKDNFY 117  
 QY 120 QHNVEKLVNVSQSLSDKSLTEKSTOYLNOHQTAAMCKWQNEGKHTEQLLESPPQTVLVP 179  
 DB 118 PHKVEKLVNVPQLNDKSPTEKSSQYLNQQTASVCKWQNEGKHAEQLLASPPACTPLP 177  
 QY 180 EGFSNANIDRSQNDHSDTDEEERDNOQFITTVKLANAKOTTEDEHAREAKSHOKCSK 239  
 DB 178 KQLSNANIQSPPTDGHSDTDEERDNOQFITPIKLANTKPTVGGQAR---SNCKCSG 234  
 QY 240 SCHPGDCCASQODEIDVVPKPSLDVSGSDVGTGSKNDKNTKIRQBSCLGNSPPFEKSE 299

```
Db 235 SRQSVKDCGQCEVVDVLPESPLSDVGAEDIGTGPKNDKLTGQSSLDGSPPFKESE 294
Qy 300 PESPMVDNKNKSCQSEADEETSPGFDQEDGSSQTANKSPRFOARADIEFRKYST 359
Db 295 PESPMVDNKNKSCQSEADEETSPVFDEQDD-RSSQTANKLSSCQAREADGLRKRYLT 353
Qy 360 KGGEVRLHFOFEGESRTGMNDLNAKLPGNISLNVECRNSKHGKDKSKITDHLMLPK 419
Db 354 KGGEVRLHFOFEGESRTGMNDLNAKLPGNISLNVECRNSKHGKDKSKITDHLMLPK 412
Qy 420 AEDRRKEQMETKQRTKIPKYPVPHLSPPKMLGTPIEMMRMPCRGIRLPLRPSAN 479
Db 413 SEDRRKEQCEVRHQRTKIPKYPVPHLSPPKMLGTPIEMMRMPCRGIRLPLRPSAN 472
Qy 480 HTVTIRVDLLRAGEVPKPFPTHYKDLWNKHVMPQSEQNLYPVEDENGERTAGSWELI 539
Db 473 HTVTIRVDLLRAGEVPKPFPTHYKDLWNKHVMPQSEQNLYPVEDENGERTAGSWELI 532
Qy 540 QTALLNKFTPNQKDALIXNVAYSXKWDFTALIDFWDKVLSEAEQAHLQYQILPDMVK 599
Db 533 QTALLNKFTPNQKDALIXNVAYSXKWDFTALIDFWDKVLSEAEQAHLQYQILPDMVK 592
Qy 600 IALCLPNICTQPIPLKQKNHSTMSQEQIASLANAFAFFCTPRRNAMKSYSSYPDI 659
Db 593 IALCLPNICTQPIPLKQKNHSTMSQEQIASLANAFAFFCTPRRNAMKSYSSYPDI 652
Qy 660 NFNRLEFGRSSRPEKLTLCFCYFRRTVEKPTGLVFTTQSLDEDPWEWERCEKPLTLH 719
Db 653 NFNRLEFGRSSRPEKLTLCFCYFRRTVEKPTGLVFTTQSLDEDPWEWERCEKPLTLH 712
Qy 720 VTYEGTTEENGQMLQVDFANRFGVGVTSAGLVQEEIRFLINPELIIISLFTVLDHNE 779
Db 713 VTYEGTTEENGQMLQVDFANRFGVGVTSAGLVQEEIRFLINPELIIISLFTVLDHNE 772
Qy 780 CLIIITGTEQSYGYAETRWRSRSHEDSGERDDCERRCTTEIVAIDALHFRYLDQVPE 839
Db 773 CLIIITGTEQSYGYAETRWRSRSHEDSGERDDCERRCTTEIVAIDALHFRYLDQVPE 832
Qy 840 KMRRELKAYCGFLRPGVPSNLSAVATGNGCGAFGDDARLKALIQILAAAAAERDVVY 899
Db 833 KVRRELKAYCGFLRPGVPSNLSAVATGNGCGAFGDDARLKALIQILAAAAAERDVVY 892
Qy 900 FTFGDSELMRDIYSMHIFLTKRLTVGD 927
Db 893 FTFGDSELMRDIYSMHIFLTKRLTVGD 920
RESULT 9
O46043 PRELIMINARY; PRT; 768 AA.
AC O46043, 1998 (TrEMBLrel. 06, Created)
DT 01-JUN-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE PARG protein.
GN PARG OR EG:114E2.1 OR CG2864.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkeley;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celnik S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan X.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballev R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
```

```
Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
Burtis J.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
Cherry J.M., Cusley S., Dahlke C., Davenport L.B., Davies P.,
de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
Dodon K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
Foster C., Gabrielian A.B., Garg N.S., Gelbart W.M., Glasser K.,
Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
Jalali M., Kalush F., Karpen G.H., Ke Z., Kemnison J.A., Ketchum K.A.,
Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M., Wang X.,
Palazzo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
Swirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195 (2000).
RN [2]
RP SEQUENCE FROM N.A.
RA Ame J.-C., Jacobson M.K.;
RT "Isolation and characterization of the cDNA encoding Drosophila
poly(ADP-ribose) glycohydrolase."
RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 46-768 FROM N.A.
RA Murphy L., Harris D., Barrell B.;
RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE OF 46-768 FROM N.A.
RA Benos P.;
RL Submitted (DEC-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AE003428; AAF45886.1; -.
DR EMBL; AF079556; AAC28734.1; -.
DR EMBL; Z98254; CAB10913.1; -.
DR FlyBase; FBgn0023216; Parg.
DR GO; GO:0016787; F:hydrolase activity; IEA.
DR InterPro; IPR007724; PARG.
DR Pfam; PF05028; PARG; 1.
KW Hydrolase.
SQ SEQUENCE 768 AA; 86288 MW; 2E620C42E643FC04 CRC64;
Query Match 20.8%; Score 1078; DB 5; Length 768;
Best Local Similarity 42.3%; Pred. No. 1.3e-68;
Matches 240; Conservative 94; Mismatches 184; Indels 50; Gaps 14;
Qy 416 RLKAEADRRKEQWETKHQRTKIPKYPVPHLSPPKMLGTPIEMMRMPCRGIRL 471
Db 45 RMGKSPDGGISSEIETEE-----PENLANSLLDSDWRGVSMEALHNRQPFLENL 94
Qy 472 PLLRPSANHTVTIRVDLLRAGEVPKPFPTHYKDLWNKHVMPQSEQNLYPVEDENGERT 531
Db 95 PPVTAGNLHRVMYQLPREF--PPRPYKSPGK--WDSEHVLPCAPSKYPRENPDGSTT 150
Qy 532 AGSRWELIQTALLNKFTPNQKDALIXNVAYSXKWDFTALIDFWDKVLSEAEQAHLQY 591
Db 151 IDFEWEMIERALLQPIKTCBELQAAIISYNTYTRDQWHFRALHQLLDSELDSETRVFE 210
Qy 592 SILPDMVKIALCLPNICTQPIPLKQKNHSTMSQEQIASLANAFAFFCTPRRNAMK 650
Db 211 DLLPRIIRLALRUPDLTIQSPVFLDKHKNASLSLSQOOISCLLANAFCLTFFRNTLKRK 270
```



```

QY 651 SEYSSYPDINFNRLPEGRSSRKPEKLTLCYCFRRV--TEKK-----PTGLVTFTRQS--L 702
Db 271 SEYTFPDINFNRLYQSTGPAVLEKLCIMHYFRVCPTRDASNVPTGVVTFVRSGLP 330
QY 703 EDFPWERCEKPL--TRLHVYEGTIEHQGQMLQVDFANRFGVGGVTSAGLVQBEIRFL 760
Db 331 BHLIDWSQSAAPLGDVPLHVDAGTIEDEGIGLLQVDFANKYLGGLVGHGVCQBEIRFV 390
QY 761 INPELIIISRLFTVLDHNECLIITGEQVSEYTYGAEYRWSRSHEDSGSRDDCERRCTE 820
Db 391 ICPELLVGKLTFCLELPPFEALVMLGAERYSNYTYGAGSFEWNGFNEDSTPRDSSGRRQTA 450
QY 821 IVAIDALHFRYLDQVPEKMRRELKAYCGFLR-----PGVSSSENLSAVATGNWGCAG 874
Db 451 IVAIDALHFAQSHQHYREDLMERELNKAYIGFVHMVTPPPG-----VATGNWGCAG 502
QY 875 FGCDARLKALIQILAAARVDVYFTFGDSLMRDIYSMHIFLTERKLTGVDVYKLLR 934
Db 503 FGDSYLKALLQMLVCAQIGRLAYTTGNVFRDDFHEMMLLFRNDGTTVOQLWS--ILR 561
QY 935 YNEECRCNSTPGP-----DIKLYPFI 956
Db 562 YSRLIKEKSKEPRENKASKKKLYDFI 589

RESULT 10
Q960N8 PRELIMINARY; PRT; 723 AA.
AC Q960N8; Q960N8;
DT 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)
DE LD42380p.
GN PARG OR EG-114E2.1 OR CG2864.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkeley;
RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,
RA Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,
RA Nunoo J., Pacleb J., Paragas V., Park S., Phouanavong S., Wan K.,
RA Yu C., Lewis S.E., Rubin G.M., Celniker S.
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY051955; AAK93379.1; -.
DR FlyBase; FBgn0023216; PARG.
DR InterPro; IPR007724; PARG.
DR Pfam; PF05028; PARG; 1.
SQ SEQUENCE 723 AA; 81109 MW; 90A5BA59775D611C CRC64;

Query Match 20.7%; Score 1073; DB 5; Length 723;
Best Local Similarity 42.2%; Pred. No. 2.6e-68;
Matches 239; Conservative 94; Mismatches 184; Indels 50; Gaps 14;

QY 417 LPKAEDRRKEQWETKHQTERKIPKVPVPHL--SPDKKMLGTPIEM--RRMPRCGIRLP 472
Db 1 MSKSPGGGISEITEEE-----PENLANSLDSWRGVSMELHNRQPFLENLP 50
QY 473 LIRPSANHTVIRVDLIRAGEVPKPTTHYKDLWNKHKVMPCEQNLYPVEDENGERTA 532
Db 51 PVTAGNLHRVMYQLPIRET--PRPYKSPGK--WDSEHVLPCAPESKYPRENPDGSTII 106
QY 533 GSRWELIQTALLNKFTROPNLKDALIKYNVAYSKKWDFTALIDFWDKVLEBAQAHLYS 592
Db 107 DFRWEMIERALLQPIKTCBELQAALISYNTYRDQWHFRALHQLDEELDESETRVFFED 166
QY 593 ILPDMVKIALCLINICTQPIPLKOKMNSITWSQIASLLANAFCTFPFRNA-KMKS 651

```

```

Db 167 LLPRIIRLALRLPOLIOFPVLLKHKXNASLSLQQOISCLIANAFCTFPRRNLKES 226
QY 652 EYSSYPDINFNRLPEGRSSRKPEKLTLCYCFRRV--TEKK-----PTGLVTFTRQS--LE 703
Db 227 EYSTFPDINFNRLYQSTGPAVLEKLCIMHYFRVCPTRDASNVPTGVVTFVRSGLPE 286
QY 704 DFPWERCEKPL--TRLHVYEGTIEHQGQMLQVDFANRFGVGGVTSAGLVQBEIRFL 761
Db 287 HLIDWSQSAAPLGDVPLHVDAGTIEDEGIGLLQVDFANKYLGGLVGHGVCQBEIRFV 346
QY 762 NPELIIISRLFTVLDHNECLIITGEQVSEYTYGAEYRWSRSHEDSGSRDDCERRCTE 821
Db 347 CPELLVGKLTFCLELPPFEALVMLGAERYSNYTYGAGSFEWNGFNEDSTPRDSSGRRQTA 406
QY 822 IVAIDALHFRYLDQVPEKMRRELKAYCGFLR-----PGVSSSENLSAVATGNWGCAG 875
Db 407 IVAIDALHFAQSHQHYREDLMERELNKAYIGFVHMVTPPPG-----VATGNWGCAG 458
QY 876 FGCDARLKALIQILAAARVDVYFTFGDSLMRDIYSMHIFLTERKLTGVDVYKLLR 935
Db 459 FGDSYLKALLQMLVCAQIGRLAYTTGNVFRDDFHEMMLLFRNDGTTVOQLWS--ILRS 517
QY 936 YNEECRCNSTPGP-----DIKLYPFI 956
Db 518 YSRLIKEKSKEPRENKASKKKLYDFI 544

RESULT 11
Q9SKB3 PRELIMINARY; PRT; 548 AA.
AC Q9SKB3; Q94ET7;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Putative poly (ADP-ribose) glycohydrolase.
GN AT2G31870 OR TEJ.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ecv. Columbia;
RA Lin X., Kaul S., Shea T.P., Fujii C.Y., Shen M., VanAken S.E.,
RA Barnstead M.E., Mason T.M., Bowman C.L., Ronning C.M., Benito M.-I.,
RA Carrera A.J., Creasy T.H., Buell C.R., Town C.D., Nierman W.C.,
RA Fraser C.M., Venter J.C.;
RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Panda S., Poirier G.G., Kay S.A.;
RT "TEJ defines a role for poly-ADP-ribosylation in establishing period
RT length of the Arabidopsis circadian oscillator.";
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC006533; AAD32285.2; -.
DR EMBL; AF394690; AAK72256.1; -.
DR PIR; B84726; B84726.
DR GO; GO:0016787; F:hydrolase activity; IEA.
DR InterPro; IPR007724; PARG.
DR Pfam; PF05028; PARG; 1.
KW Hydrolase.
SQ SEQUENCE 548 AA; 62169 MW; FIA79FDA157C3329 CRC64;

Query Match 11.2%; Score 579; DB 10; Length 548;
Best Local Similarity 32.4%; Pred. No. 5.3e-33;
Matches 146; Conservative 64; Mismatches 137; Indels 104; Gaps 10;

QY 577 WDKVLEBAQAHLYSILPDMVKIALCLINICTQPIPLKOKMNS 622
Db 79 FDELIDKESKRWDFEIPALASLLQFPFLSEVHFQADNVISGKTLRLNSQAGI 138
QY 623 ITWSQIASLLANAFCTFPFRNAKMSEYSSYPDINFNLIFE----GRSSRKPEKLT 678

```

```

Db 139 VFLSQBLIGALLACSPFLPPDDNRGAK-----HLPVINFDFHSLFASLYISYSQSQESKIRC 194
QY 679 LFCYFRRVTEKKGTGLVTRQ--SLEDPEWERCCKPLTRHLVTVTEGTEENGQGMLOV 736
Db 195 IMHYFRFCVPIGVISFERKTAAPDADFWSKSDVSLCAFKVHSGFLIEDQDVALEV 254
QY 737 DFANRVGGVTSAGLVQOEIRFLINPELIISRLFTVLDHNECLITGTEQYSEYTGVA 796
Db 255 DFANKYLGGSLSRGCVQEBIRFMWNPINPELIAGMLFLPRMDNEAIBVGAERFSCVTGVA 314
QY 797 ETRWRSHEDGSDRDCERRCTEIVAI-----HFRYLDQFVPEKMRRLNKAIC 850
Db 315 SSFPAAGEYIDKKAMPDFKERRTRIVAI-----LLEINKALC 367
QY 851 GFLR----- 854
Db 368 GFLNCKAWEHQIFWDEGNEIQLVNRGRDGLRTEITATSHRTPLNDVEMREKPAANN 427
QY 855 -----PGVSSENL--SAVATNGWCGAGFGDARLKALIQILAAAAAERDVV-YTTFGD 904
Db 428 LIRDFYVEGVNEDHEDDGVATNGWCGVGGDPPELKATIQWLAASTQTRPFISYITFG- 486
QY 905 SELMRDIYSMHIFLTERKLTGVGVYKLLRY 935
Db 487 VEALRNLDQVTKWILSHKWTGVGLNMMLEY 517

```

## RESULT 12

```

ID Q8VYAI PRELIMINARY; PRT; 522 AA.
AC Q8VYAI;
DT 01-MAR-2002 (Tremblrel. 20, Created)
DT 01-MAR-2002 (Tremblrel. 20, Last sequence update)
DE 01-JUN-2003 (Tremblrel. 24, Last annotation update)
DE Putative poly(ADP-ribose) glycohydrolase.
GN AT2G31870.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosid II; Brassicales; Brassicaceae; Arabidopsi.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Nguyen M., Karlin-Neumann G., Southwick A., Lam B., Miranda M.,
RA Palm C.J., Bowser L., Jones T., Banh J., Carninci P., Chen H.,
RA Cheuk R., Chung M.K., Hayashizaki Y., Ishida J., Kamiya A., Kawai J.,
RA Kim C., Lin J., Liu S.X., Narusaka M., Pham P.K., Sakano H.,
RA Sakurai T., Satou M., Seki M., Shinn P., Yamada K., Shinozaki K.,
RA Ecker J., Theologis A., Davis R.W.;
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY072330; AAL61937.1; -
DR GO; GO:0016787; F-hydrolase activity; IEA.
DR InterPro; IPR007724; PARG.
DR Pfam; PF05028; PARG; 1.
KW Hydrolase.
SQ SEQUENCE 522 AA; 58942 MW; 6FF98685D867CA15 CRC64;

```

```

Query Match 10.48; Score 539; DB 10; Length 522;
Best Local Similarity 30.74; Pred. No. 3.7e-30;
Matches 133; Conservative 82; Mismatches 136; Indels 80; Gaps 11;
QY 576 FWDVLBEAEAHLYQSILPMWIKALCLPNICTQPIPLKQKNH----- 621
Db 80 FPDKKISREANFGEVWPALCRLLQLPSMLEKH----QKADHVLGDKSGRLILGP 135
QY 622 ----SITMSQBIASLIANAFCTFPRNARKSESYDINRLE-----GRSRKP 673
Db 136 QEAGVILLQELIAALLACSPFLFPEVDRLK----NLOGINFGSLFSPFMRHCKQE 191
QY 674 EKLKTLFCYFRVTEKPTGLVTRQSL-----DPEP---WERCEKPLRLHVTY 722
Db 192 NKIKRLIHYFGRICRMWPTGFVSFERKILPLEYHPHFVSYPKADSWANSVTPLCSEIHT 251

```

```

QY 723 EGTIENGGMQLQVDFANFVGGVTSAGLVQOEIRFLINPELIISRLFTVLDHNECLI 782
Db 252 SGAIEDPOPCALEVDPADEYFGLTSLYDTLOEIRFVINPELIAGMIFLPRMDANEAE 311
QY 783 ITCTEYQSEYTGVAETFRWRSRSHEDGSDRDCERRCTEIVAI-----HFRYLDQFVPEKMR 842
Db 312 IGVVERFSYTGYSFGSFAGDYTDNKDLDFRRKTRVIAIDAMP-DPGMGQYKLDALI 370
QY 843 RELNKAYCGFLR-----PGVSS-----ENLSAVA 866
Db 371 REVNKAFSGYHCKYKNIDVKGHPPEASSHVPLTSDSASQVIESSHRWCIDHEEKIGVA 430
QY 867 TGNWCGAGFGDARLKALIQILAAAAAERDVV-YTTFGSELMRDIYSMHIFLTERKLT 925
Db 431 TGNWCGVGGDPPELKIMQLWLAISQGRPFMSYITFG-LQALONLQNVIEVALQEMTV 489
QY 926 GDVYKLLRYNE 938
Db 490 GDLWKKLVEYSSE 502

```

## RESULT 13

```

ID Q867XO PRELIMINARY; PRT; 781 AA.
AC Q867XO;
DT 01-JUN-2003 (Tremblrel. 24, Created)
DT 01-JUN-2003 (Tremblrel. 24, Last sequence update)
DE 01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE Poly ADP-ribose metabolism enzyme-3 long form (C. elegans PME-3
DE protein) (Corresponding sequence F20C5.1a).
GN F20C5.1 OR PME-3.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Gagnon S.N., Hardy I., Desnoyers S.;
RT "Characterization of poly(ADP-ribose) glycohydrolases in the nematode
RT Caenorhabditis elegans.";
RL Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=99069613; PubMed=9851916;
RA none;
RT "Genome sequence of the nematode C.elegans: A platform for
RT investigating biology.";
RL Science 282:2012-2018(1998).
RN [3]
RP SEQUENCE FROM N.A.
RA Matthews P.;
RL Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY185493; AAO26316.1; -
DR EMBL; Z68161; CAD89735.1; -
DR Wormpep; F20C5.1a; CE33775.
DR InterPro; IPR007724; PARG.
DR Pfam; PF05028; PARG; 1.
SQ SEQUENCE 781 AA; 89273 MW; 30447D06605CED0A CRC64;

```

```

Query Match 9.38; Score 480.5; DB 5; Length 781;
Best Local Similarity 24.48; Pred. No. 1.1e-25;
Matches 215; Conservative 136; Mismatches 310; Indels 219; Gaps 39;
QY 145 YLNOHQTAAACKWQNEKHTQELLESEPQVTILVPEQFSNANIDRSQNDHSDTSDSEN 204
Db 27 FAHQVPMKRLKTEHNTTESKD-----PEEPKSRDVFVSSQSDSEQSDAEN 77
QY 205 -----RDNQQLT-TVKLANAQ-----TTEDEHAREAKSHOKCKSHCPGDCASQ 252
Db 78 PEIAKEVSENCENTETLKISNIESLDNVTERSEHTLD--NH-----KSTPEMEE----- 125
QY 253 DEIDVPEKPLSDVGSVDGTGSKNDKLIHQESCLGNSPFPEKESEPESPMDVDNKGNS 312

```

Db 126 ---DYNKNSNI-----DVAINSEDEDELVEEN-----NKE-----WEDGEQV 160  
QY 313 QDSEADEE---TSPGDEQE-----DGSSSQATANKPSRFQARDADIEFRKRYSTKGE 363  
Db 161 QODLFADDOQLIEYPGIMKDTTQDIDTSDSEVETAKMEMIEETAD-----STFVGE 213  
QY 364 VRLHFQFGGSSRTGMNDLNAKLPNISLNVCEKNSQKHGKOSKITDHLMLPKAEDR 423  
Db 214 DSKNR-----QSGTTSDEVDADSQINLATKTVRTSS-----SFLSTVSTSCAP 258  
QY 424 RKBQWETHQRTKIPKVPPLHS--PDKKWLGTPIEMRMPCGIRLPLLRSAHNT 481  
Db 259 AKGRARMYQKEKGVIAFTGNTLQPD-----LNKVPDRNY- 297  
QY 482 VTIIVDLRAGVPPKPPHYKDL--WDNKH---VRMPCSEQLNYPVEDENGERTAGSRWE 537  
Db 298 -----RYCTIPN-FPASQKLRNDRYGPKIVLP-----QW- 328  
QY 538 LIQTALLNKFTPPQNLKAILKYN---VAYSKKWDFTA-----LIDFWD---KVLBEA 584  
Db 329 -----REFDGRGRRRDSYFFKFKLDGILKCYKTTGYFMFVGLLHNMWFEFDPDITYKL 381  
QY 585 EAQHLVQSILPMV-----KIALCLPNICTQPIPLKQKNNHSITMSQEQIA 631  
Db 382 PALEMYKEMSELVGRREEVLEKFAVARIAKTAEDLLPERIYRLVGDV-ESATLSHKQCA 440  
QY 632 SLLANAFCTFPRRNAMKSEYSSYPDINFNRLEFGRSSRKPEKLTLCFYFRRYTEKKP 691  
Db 441 ALVARMFF-----ARPOSPFS-----FCRLSSDKSICVEKUKFLFTYFDKMSMDPP 487  
QY 692 TGLVTF--TROSLDF-PEWERCXPLRLHVT--YEGTIBENGQMLQVDFANRVVGG 746  
Db 488 DGAVSPLTKMDKDTNEEWK--DKKLSLPEVEFFDEMLIEDTALCTQVDFAHEHLGG 545  
QY 747 VTSAGLVQBEIRLPLNPELIISELFTVLDHNECLIIITGEQSEVTVGAETVRW----- 801  
Db 546 VLNHGVSQBEIRFIMCPGEMVMVGMMLCEKMKQLEAISIVGAYVSSYTGHTLKWAELOP 605  
QY 802 SRSHEDGSE--RDCERRCTEIVADALHFR-----YLDQFVPEKMRRLNKAYCOFLRP 855  
Db 606 NBSRQNTNFRDFRGLRVETIADAILFKGSKLDQTEQLNANKANIIRMKKASIGFMSQ 665  
QY 856 GVSSENLSAVATNGWCGAFGGQARUKALIQIIIAAAERDVYFFGSELMDIYSMH 915  
Db 666 GPKFTMI-PIVTCWVGCGAFNGDKPLKFLIIQVIAAGVADRLHFCFSFGEPELAACKKII 724  
QY 916 IFLTERKLTVDGVYKULLRYNEECRNCSTPGPDIKLYPF 955  
Db 725 ERMKQKDVTLGMLFSMI-----NNTGLPHKHFEFYVF 756  
  
RESULT 14  
Q19637  
ID Q19637 PRELIMINARY; PRT; 764 AA.  
AC Q19637;  
DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
DT 01-MAR-2003 (TrEMBLrel. 23, last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, last annotation update)  
DE PME-3 protein (corresponding sequence F20C5.1) (Poly ADP-ribose  
DE metabolism enzyme-3 short form).  
GN F20C5.1 OR PME-3.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidae;  
OC Rhabditidae; Peloderinae; Caenorhabditis.  
OX NCBI\_TaxID=6239;  
RN [1]  
RN SEQUENCE FROM N.A.  
RP Matthews P.;  
RA Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.  
RL [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=99069613; PubMed=9851916;

RA none;  
RT "Genome sequence of the nematode C.elegans: A platform for  
RT investigating biology";  
RL Science 282:2012-2018 (1998).  
RN [3]  
RP SEQUENCE FROM N.A.  
RA Gagnon S.N., Hardy I., Deanoyers S.;  
RT "Characterization of poly(ADP-ribose) glycohydrolases in the nematode  
RT Caenorhabditis elegans";  
RL Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL; 268161; CRA92299.2; -;  
DR EMBL; AY185494; AAO26317.1; -;  
DR PIR; T21138; T21138.  
DR WormPep; F20C5.1b; CE32867.  
DR InterPro; IPR007724; PARG.  
DR Pfam; PF05028; PARG; 1.  
SQ SEQUENCE 764 AA; 87441 MW; B391703B1D0CE79A CRC64;  
  
Query Match 9.1%; Score 473; DB 5; Length 764;  
Best Local Similarity 23.6%; Pred. No. 3.5e-25;  
Matches 211; Conservative 128; Mismatches 290; Indels 266; Gaps 38;  
  
QY 145 YLNQHTQAMCKWQEGKHTEQLLESEFQTVTLVPEQFSNANIDRSFONDHSDHSDTSDSEN 204  
Db 27 FAHQVPTMKRRKRLTEHGNTEKED-----PEPKSRDVFVSSSDSESQEDSAEN 77  
QY 205 -----RDNQOFLT-TVKLANAKQ-----TTDEHAREAKSHQKSCHPGEDCASCQQ 252  
Db 78 PEIAKEVSENCENLTETLIKINIESLDNVTSEHTLD--NH-----KSTEPWEE----- 125  
QY 253 DEIDVVPKSPISDVGSSEVDGTGSKNDKLIHQESCLGNSPPPEKSESESPESPMVDVNSKNS 312  
Db 126 -----DYNKNSNI-----DVAINSEDEDELVEEN-----NKEMR-----DGEVQVQL 163  
QY 313 QDSEADEE---TSPGDEQE-----DGSSSQATANKPSRFQARDADIEF-----RKR 356  
Db 164 SODLFADDOQLIEYPGIMKDTTQDIDTSDSEVETAKMEMIEETADSTFVGEDSKATYK 223  
QY 357 YSTKG-----GEVRLH-----FQEGGESRTGMNDLNAKLP----- 388  
Db 224 VRTSSSFLSTVSTCEAPAKGARMYQKEKXHVIAFTG-NITLPDLNKVDPPDRNYR 282  
QY 389 -----NSSLNVCEKNSQKHGKOSKITDHLMLPKAEDRRKEQWETHORTE-----BK 438  
Db 283 CTIPNPASQKGLREDNRYGPK-----IVLP-----QWRFEFDSGRRRDSYFFKRRK 330  
QY 439 IPKYVPPHLSPKKMLGTPIEMRM-PRCGIRLPLLRPSANHTVIRVDLLRAGEVVPK 497  
Db 331 LDGILKCYKTYGYFMFVGLLHNMWFEFDPDITYKLPALE----- 368  
QY 498 FPTHYKDLWMDKHWKMPFCSEQLNYPVEDENGERTAGSRWELI-OTALLNKFTPPQNLKDA 556  
Db 369 --MYKEM-----SELVGRREEVLEKFAVARIA-- 394  
QY 557 ILKYNVAYSKKMDFTALIDFWDKVLBEAQAHLVQSILPDMVKIALCLPNICTQPIPLLK 616  
Db 395 -----KTAEDILPERIYR-LVGDV----- 412  
QY 617 QKNNHSITMSQEQIASILANAFCTFPRRNAMKSEYSSYPDINFNRLEFGRSSRKPEKL 676  
Db 413 -----ESATLSHKQCAALVARMFF-----ARPDSPFS-----FCRLSSDKSICVEKL 455  
QY 677 KTLFCYFRVRVTEKKPTGLVTF--TROSLDF-PEWERCXPLRLHVT--YEGTIBENGQ 731  
Db 456 KFLFTYFDKMSMDPPDGAVSFRITKMDKDTNEEWK--DKKLSLPEVEFFDEMLIEDTA 513  
QY 732 GMLQVDFANRVVGGVTSAGLVQBEIRLPLNPELIISELFTVLDHNECLIIITGEQYSE 791  
Db 514 LCTQVDFAHEHLGGVGLNHGVSQBEIRFIMCPGEMVMVGMMLCEKMKQLEAISIVGAYVSS 573  
QY 792 YTGVAETVRW-----SRSHEDGSE--RDCERRCTEIVADALHFR-----YLDQFVPEK 840  
Db 574 YTYGHTLKWAELOPNSRQNTNFRDFRGLRVETIADAILFKGSKLDQTEQLNANKAN 633

QY 688 EKKPTGLVTF-----RSLRDEPEWE-RCEKPLRLHVTVEGTIEENGQGMLOVDFA 740  
Db 201 IDPPIGAVSFRKWRTHKQYLEN---WKLEFNLLPDVQVFDKMSLEETAL-CTQIDFA 256  
QY 741 RFVGGVTSAGLVQBEIRF-INPELIISRLFTVELDHNELIITGTQVSEYTGVAETFR 800  
Db 257 KRIGGGVLKGGAVQBEIRFMPCPMVAILLNDVTQDLEAISIVGAYVFSSTGYSTNLK 316  
QY 801 WSR-----SHEDGSRDDCERRCTEIVADALH-----FRRLDQFVPEKMERELNKAY 849  
Db 317 WAKITPKHSAQNNSFRDQFRLQETETVAIDAVRNAGTPECLNQLTTEKLTREVRKAA 376  
QY 850 CGFLRPGVSSSNNLSAVATGNWGGAGFGDARLKALIQILAAAAAARDVVYFTFGDSELMR 909  
Db 377 IGFLSAGDGFSKI-PVVGWGGGAGFRGNKPLKFLQIVACGISDRPLQCFCTFGDTLAK 435  
QY 910 DIYSMEHIFLTERKLTVDVYKLL 932  
Db 436 KCEMMTLFRNNNVRTGOLFLLI 458

Search completed: May 26, 2004, 18:45:59  
Job time : 81.8912 secs

QY 841 MRRELKAYCGFLRPGVSSSNNLSAVATGNWGGAGFGDARLKALIQILAAAAAARDVVYF 900  
Db 634 IIREMKASIGFMSQGFKEFNI-PIVTGWWGGAGFNGDKPLKFLIIQIVIAAGVADRLHFC 692  
QY 901 TFGDSLMRDIYSMHIFLTERKLTVDVYKLLRYNECRNCSTGPDIKLYPF 955  
Db 693 SFGPELAAKCKKIIERMKQKVTLGMLFSMI-----NNTGLPHKHFEFYVF 739

RESULT 15

Q9NSL4 PRELIMINARY; PRT; 485 AA.  
AC Q9NSL4;  
DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
DE Hypothetical protein H23L24.5 (Poly ADP-ribose metabolism enzyme  
4).  
DE H23L24.5 OR PME-4.  
GN H23L24.5 OR PME-4.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
OC Rhabditidae; Peloderinae; Caenorhabditis.  
OX NCBI\_TaxID=6239;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Bristol N2;  
RX MEDLINE=99069613; PubMed=9851916;  
RA Waterston R.;  
RT "Genome sequence of the nematode C. elegans: a platform for  
investigating biology. The C. elegans Sequencing Consortium.";  
RL Science 282:2012-2018 (1998).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Bristol N2;  
RA Abu-Threideh J., Lehnert L.;  
RT "The sequence of C. elegans cosmid H23L24.";  
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Bristol N2;  
RA Waterston R.;  
RL Submitted (DEC-2002) to the EMBL/GenBank/DBJ databases.  
RN [4]  
RP SEQUENCE FROM N.A.  
RA Gagnon S.N., Hardy I., Desnoyers S.;  
RT "Molecular characterization of poly ADP-ribose glycohydrolases in  
Caenorhabditis elegans.";  
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AC006662; AAF39896.2; -  
DR EMBL; AF548468; AAN40699.1; -  
DR WormPep; H23L24.5; CE32685.  
DR InterPro; IPR007724; PARG.  
DR Pfam; PF05028; PARG; 1.  
KW Hypothetical protein.  
SQ SEQUENCE 485 AA; 55574 MW; 86367DDBAB15081F CRC64;

Query Match 8.7%; Score 453; DB 5; Length 485;  
Best Local Similarity 30.7%; Pred. No. 5e-24;  
Matches 136; Conservative 74; Mismatches 157; Indels 76; Gaps 15;

QY 536 WELIQTALLNKFR-PQNLKDALIK--YNAVYS-----KKWDFALIDF 576  
Db 46 FELLETGVSQQNRCDQNLFNELKTKYNGGYSQFEDLLFKIWGSEKERFDLPALKSF 105  
QY 577 WDKVLE---EBAQHLQSIPLDMVKIALCLPNICTQPIPLKKQKN-----HSITMSQ 627  
Db 106 YRKMSIIVGEDE-----VLEKLRIVITYKSAC-----EVLPEKIYELVGDIESATFSH 154  
QY 628 EQIASLLANAFCTPPRNNAKMKSEYSSYPDINRNLFEGRSSRKPEKLTLCYFRVRT 687  
Db 155 IQCASLIAMWFFSDTPR-----LSFIIILQKTCVAVKELFLTYFDKMS 200

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 26, 2004, 18:18:25 ; Search time 77.971 Seconds  
(without alignments)  
3953.539 Million cell updates/sec

Title: US-09-302-812-2  
Perfect score: 5184  
Sequence: 1 MSAGPGCEPCTKRPWDAAA.....YHAVESCTQTNNQPCQTGA 977

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SPTREMBL\_25.\*

- 1: sp\_archaea.\*
- 2: sp\_bacteria.\*
- 3: sp\_fungi.\*
- 4: sp\_human.\*
- 5: sp\_invertebrate.\*
- 6: sp\_mammal.\*
- 7: sp\_mhc.\*
- 8: sp\_organelle.\*
- 9: sp\_phase.\*
- 10: sp\_plant.\*
- 11: sp\_rodent.\*
- 12: sp\_virus.\*
- 13: sp\_vertebrate.\*
- 14: sp\_unclassified.\*
- 15: sp\_rvirus.\*
- 16: sp\_bacteriap.\*
- 17: sp\_archaeap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	5184	100.0	977	6	O02776	O02776 bos taurus
2	4634.5	90.6	976	4	O86W56	O86W56 homo sapien
3	4689.5	90.5	976	4	Q72742	Q72742 homo sapien
4	4635.5	89.4	976	4	Q9Y4W7	Q9Y4W7 homo sapien
5	4332.5	83.6	968	11	O88622	O88622 mus musculus
6	4329	83.5	961	11	O80YQ6	O80YQ6 mus musculus
7	4295.5	82.9	972	11	Q9QYM2	Q9QYM2 rattus norv
8	4125	79.6	920	11	O8CB72	O8CB72 mus musculus
9	1063.5	20.5	768	5	O46043	O46043 drosophila
10	1061.5	20.5	723	5	Q96ON8	Q96ON8 drosophila
11	585	11.3	548	10	Q9SKB3	Q9SKB3 arabidopsis
12	533	10.3	522	10	Q8VY41	Q8VY41 arabidopsis
13	481.5	9.3	781	5	O867X0	O867X0 caenorhabdi
14	475	9.2	764	5	Q19637	Q19637 caenorhabdi
15	472	9.1	485	5	Q9N5L4	Q9N5L4 caenorhabdi
16	335.5	6.5	368	5	Q86G14	Q86G14 toxoplasma

17	331	6.4	364	10	Q9SKB4	Q9SKB4 arabidopsis
18	185	3.6	508	12	Q9E234	Q9E234 helicoverpa
19	181	3.5	571	5	Q8NTN7	Q8NTN7 trichinella
20	180	3.5	510	12	Q99GU9	Q99GU9 helicoverpa
21	180	3.5	2081	10	Q9LH98	Q9LH98 arabidopsis
22	179.5	3.5	1078	5	Q963T1	Q963T1 plasmodium
23	171.5	3.3	1444	5	Q9VTN2	Q9VTN2 drosophila
24	171.5	3.3	1514	5	O8SY55	O8SY55 drosophila
25	171	3.3	791	13	Q9DGL1	Q9DGL1 fugu rubrip
26	169.5	3.3	678	13	Q803F8	O803F8 brachydanio
27	167	3.2	455	5	Q86A18	Q86A18 dictyosteli
28	165	3.2	968	5	Q8I5W9	Q8I5W9 plasmodium
29	164.5	3.2	1012	5	Q8SXP2	Q8SXP2 drosophila
30	164.5	3.2	1330	5	Q9VJ87	Q9VJ87 drosophila
31	164	3.2	1115	5	Q17863	Q17863 caenorhabdi
32	164	3.2	1618	16	Q7UYM1	Q7UYM1 rhodopirell
33	163.5	3.2	498	5	Q8NTN8	Q8NTN8 trichinella
34	162.5	3.1	1071	10	Q9M0D9	Q9M0D9 arabidopsis
35	162.5	3.1	7210	5	Q9V7G8	Q9V7G8 drosophila
36	162.5	3.1	9270	5	Q8MLD9	Q8MLD9 drosophila
37	161.5	3.1	3398	13	O8QFV6	O8QFV6 gallus gall
38	161	3.1	691	4	Q96QF7	Q96QF7 homo sapien
39	160.5	3.1	1152	4	Q92603	Q92603 homo sapien
40	160.5	3.1	1204	4	O00211	O00211 homo sapien
41	159.5	3.1	1156	5	Q8IM05	Q8IM05 plasmodium
42	159.5	3.1	3111	5	Q9VH10	Q9VH10 drosophila
43	159	3.1	1535	10	Q9LXZ9	Q9LXZ9 arabidopsis
44	159	3.1	1817	4	O43166	O43166 homo sapien
45	159	3.1	8081	5	Q7Z120	Q7Z120 caenorhabdi

ALIGNMENTS

RESULT 1

O02776	PRELIMINARY;	PRT;	977 AA.
AC	O02776	PRELIMINARY;	PRT;
DT	O02776;	(TREMREL. 04, Created)	
DT	01-JUL-1997	(TREMREL. 04, Last sequence update)	
DT	01-JUN-1997	(TREMREL. 04, Last sequence update)	
DT	01-JUN-2003	(TREMREL. 24, Last annotation update)	
DE	Poly(ADP-ribose)	glycohydrolase.	
GN	BPARG.		
OS	Bos taurus (Bovine).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;		
OC	Bovidae; Bovinae; Bos.		
OC	NCBI_TaxID=9913;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=92727328; PubMed=9115250;		
RA	Lin W., Ame J.C., Aboul-Ela N., Jacobson E.L., Jacobson M.K.;		
RT	"Isolation and characterization of the cDNA encoding bovine poly(ADP-ribose) glycohydrolase."		
RL	J. Biol. Chem. 272:11895-11901(1997).		
DR	EMBL: U78975; AAB53370.1; --		
DR	GO; GO:0016787; F:hydrolase activity; IEA.		
DR	InterPro; IPR007724; PARG.		
DR	Pfam; PF05028; PARG; 1.		
KW	Hydrolase.		
SQ	SEQUENCE	977 AA; 110837 MW; 87D2100F979DF377 CRC64;	

Query Match 100.0%; Score 5184; DB 6; Length 977;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 977; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy	1	MSAGPGCEPCTKRPWDAAATSPPAASDARSFPGQRRLVDSKDAPVQFRVPPSSGCA	60
Db	1	MSAGPGCEPCTKRPWDAAATSPPAASDARSFPGQRRLVDSKDAPVQFRVPPSSGCA	60
Oy	61	GRAGQHRGATSILVFKQKTTISWMDTKGKIVSESLHSENNNTREESMMSSVQKKNFY	120
Db	61	GRAGQHRGATSILVFKQKTTISWMDTKGKIVSESLHSENNNTREESMMSSVQKKNFY	120

```

QY 121 QHNMEKLENSQLGFDKSPVKGTOYLKQHOTAAKCKWQNEGPHSERLLESEPPAVTILVP 180
Db 121 QHNMEKLENSQLGFDKSPVKGTOYLKQHOTAAKCKWQNEGPHSERLLESEPPAVTILVP 180
QY 181 EQFSNANVDQSPKDDHSDTNSSESRDQQLFTHVKLANAKQWMEDEQGRARSHQKCGK 240
Db 181 EQFSNANVDQSPKDDHSDTNSSESRDQQLFTHVKLANAKQWMEDEQGRARSHQKCGK 240
QY 241 ACHPAEACAGCQOEETDVVSEPLSDTSESDVGTGLKNANRLNROESSLGNSSPPPEKSE 300
Db 241 ACHPAEACAGCQOEETDVVSEPLSDTSESDVGTGLKNANRLNROESSLGNSSPPPEKSE 300
QY 301 PESPMVDVNSKNSCODSEADETSPGFDEQEDSSSAQTANKPSRFPQPREADTELAKRSSA 360
Db 301 PESPMVDVNSKNSCODSEADETSPGFDEQEDSSSAQTANKPSRFPQPREADTELAKRSSA 360
QY 361 KGGEIRLHFQFEGGESRAGMNDVNAKPGSTSSLVNVECRNSQHGKDKSKI TDHFMVVK 420
Db 361 KGGEIRLHFQFEGGESRAGMNDVNAKPGSTSSLVNVECRNSQHGKDKSKI TDHFMVVK 420
QY 421 AEDKKEQCEMKHQRTERKIPKYPHLSPPDKKWLGTPIEEMRMPCGIRLPPPLRPSAN 480
Db 421 AEDKKEQCEMKHQRTERKIPKYPHLSPPDKKWLGTPIEEMRMPCGIRLPPPLRPSAN 480
QY 481 HTVTIRVDLLRIGEVKPPPTTHFKDILNDKHKVMPCEQNLVPEDENGERAAGSRWELI 540
Db 481 HTVTIRVDLLRIGEVKPPPTTHFKDILNDKHKVMPCEQNLVPEDENGERAAGSRWELI 540
QY 541 QPALLNRLTRPQNLKDAILKYNVAYSKKWDFTALIDFWDKVLFEAAQHLQYSLIPDMVK 600
Db 541 QPALLNRLTRPQNLKDAILKYNVAYSKKWDFTALIDFWDKVLFEAAQHLQYSLIPDMVK 600
QY 601 IALCLPNICTOPIPLKQKNHSITWSQEQIASLLANAFCTPPRNNAKMSSEYSPDI 660
Db 601 IALCLPNICTOPIPLKQKNHSITWSQEQIASLLANAFCTPPRNNAKMSSEYSPDI 660
QY 661 NFNRLEFEGSSSRKPEKLTLCYFVRVTEKPTGLVTFTRQSLDPPWERCEKLTLLH 720
Db 661 NFNRLEFEGSSSRKPEKLTLCYFVRVTEKPTGLVTFTRQSLDPPWERCEKLTLLH 720
QY 721 VTYEGTIEGNGQGLQVDFANFRVGGVTSAGLVQBEIRFLINPELIVSELFTVLDHNE 780
Db 721 VTYEGTIEGNGQGLQVDFANFRVGGVTSAGLVQBEIRFLINPELIVSELFTVLDHNE 780
QY 781 CLIIITGEQSEYTGVAETVYRWARSHEDSRDDWQRTTEIIVADALHFRVYLDQFVPE 840
Db 781 CLIIITGEQSEYTGVAETVYRWARSHEDSRDDWQRTTEIIVADALHFRVYLDQFVPE 840
QY 841 KIRRELKAYCGFLRPGVSSENLSAVATGNWCGAFGGDARLXALI QILAAA VARDVY 900
Db 841 KIRRELKAYCGFLRPGVSSENLSAVATGNWCGAFGGDARLXALI QILAAA VARDVY 900
QY 901 FTGDSLMEDDIYSMTFTLTERKLTGVEVVKLLIRYNECRNCSPTGPDILKLYPIYHA 960
Db 901 FTGDSLMEDDIYSMTFTLTERKLTGVEVVKLLIRYNECRNCSPTGPDILKLYPIYHA 960
QY 961 VESCTQTTNPGORTGA 977
Db 961 VESCTQTTNPGORTGA 977

```

RESULT 2

```

Q86W56 PRELIMINARY; PRT; 976 AA.
ID Q86W56
AC Q86W56;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Similar to poly (ADP-ribose) glycohydrolase.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

```

```

OK NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Skin;
RA Strausberg R.;
RL Submitted (APR-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC050560; AAH50560.1;
DR GO; GO:0016787; F:hydrolase activity; IEA.
DR InterPro; IPR007724; PARG.
DR Pfam; PF05028; PARG; 1.
KW Hydrolase.
SQ SEQUENCE 976 AA; 111109 MW; D6646353C6D0180E CRC64;

```

Query Match 90.6%; Score 4694.5; DB 4; Length 976;  
 Best Local Similarity 90.5%; Pred. No. 0;  
 Matches 883; Conservative 29; Mismatches 63; Indels 1; Gaps 1;

```

QY 1 MSAGFGCEPCTKRPRWDAATSPPAASDARFPQRQRRVLDSDAPVFRVPPSSGAL 60
Db 1 MNAGFGCEPCTKRPRWGAATTS-PAASDARFPQRQRRVLDPKDAHVQFRVPPSPACVP 59
QY 61 GRAGQHRGSATSLVFKQKTIITSMWDTGKITVESLSLHSENNNTREESMSSVQKDNFY 120
Db 60 GRAGQHRGSATSLVFKQKTIITSMWDTGKITAESLSDSKENNNTRIIESMSSVQKDNFY 119
QY 121 QHNMEKLENSQLGFDKSPVKGTOYLKQHOTAAKCKWQNEGPHSERLLESEPPAVTILVP 180
Db 120 QHNMEKLENSQLGFDKSPVKGTOYLKQHOTAAKCKWQNEGPHSERLLESEPPAVTILVP 179
QY 181 EQFSNANVDQSPKDDHSDTNSSESRDQQLFTHVKLANAKQWMEDEQGRARSHQKCGK 240
Db 180 EQFSNANVDQSPKDDHSDTNSSESRDQQLFTHVKLANAKQWMEDEQGRARSHQKCGK 239
QY 241 ACHPAEACAGCQOEETDVVSEPLSDTSESDVGTGLKNANRLNROESSLGNSSPPPEKSE 300
Db 240 SCDPGDCASCQOEIDVVPESPLSDVSGEDVGTGPKNDKLTROESCIGNSPPEKSE 299
QY 301 PESPMVDVNSKNSCODSEADETSPGFDEQEDSSSAQTANKPSRFPQPREADTELAKRSSA 360
Db 300 PESPMVDVNSKNSCODSEADETSPGFDEQEDSSSAQTANKPSRFPQPREADTELAKRSSA 359
QY 361 KGGEIRLHFQFEGGESRAGMNDVNAKPGSTSSLVNVECRNSQHGKDKSKI TDHFMVVK 420
Db 360 KGGEIRLHFQFEGGESRAGMNDVNAKPGSTSSLVNVECRNSQHGKDKSKI TDHFMVVK 419
QY 421 AEDKKEQCEMKHQRTERKIPKYPHLSPPDKKWLGTPIEEMRMPCGIRLPPPLRPSAN 480
Db 420 AEDRKRQEWETKHQRTERKIPKYPHLSPPDKKWLGTPIEEMRMPCGIRLPPPLRPSAN 479
QY 481 HTVTIRVDLLRIGEVKPPPTTHFKDILNDKHKVMPCEQNLVPEDENGERAAGSRWELI 540
Db 480 HTVTIRVDLLRIGEVKPPPTTHFKDILNDKHKVMPCEQNLVPEDENGERTAGSRWELI 539
QY 541 QPALLNRLTRPQNLKDAILKYNVAYSKKWDFTALIDFWDKVLFEAAQHLQYSLIPDMVK 600
Db 540 QPALLNRLTRPQNLKDAILKYNVAYSKKWDFTALIDFWDKVLFEAAQHLQYSLIPDMVK 599
QY 601 IALCLPNICTOPIPLKQKNHSITWSQEQIASLLANAFCTPPRNNAKMSSEYSPDI 660
Db 600 IALCLPNICTOPIPLKQKNHSITWSQEQIASLLANAFCTPPRNNAKMSSEYSPDI 659
QY 661 NFNRLEFEGSSSRKPEKLTLCYFVRVTEKPTGLVTFTRQSLDPPWERCEKLTLLH 720
Db 660 NFNRLEFEGSSSRKPEKLTLCYFVRVTEKPTGLVTFTRQSLDPPWERCEKLTLLH 719
QY 721 VTYEGTIEGNGQGLQVDFANFRVGGVTSAGLVQBEIRFLINPELIVSELFTVLDHNE 780
Db 720 VTYEGTIEGNGQGLQVDFANFRVGGVTSAGLVQBEIRFLINPELIVSELFTVLDHNE 779
QY 781 CLIIITGEQSEYTGVAETVYRWARSHEDSRDDWQRTTEIIVADALHFRVYLDQFVPE 840
Db 780 CLIIITGEQSEYTGVAETVYRWARSHEDSRDDWQRTTEIIVADALHFRVYLDQFVPE 839

```

QY 841 KIRRELKAYCGFLRPGVSSSENLSAVATGNWCGAFGGDARLKALIQILAAVAERDVVY 900  
 DB 840 KMRRELKAYCGFLRPGVSSSENLSAVATGNWCGAFGGDARLKALIQILAAVAERDVVY 899  
 QY 901 FTFGSELMDIYSMTFLTERKLTGVEVYKLLRYNEECRCNSTPGDILKYPFIYHA 960  
 DB 900 FTFGSELMDIYSMTFLTERKLTGVEVYKLLRYNEECRCNSTPGDILKYPFIYHA 959  
 QY 961 VESCTQTTNPGQRTG 976  
 DB 960 VESCAETADHSQRTG 975  
 RESULT 3  
 Q7Z742 PRELIMINARY; PRT; 976 AA.  
 AC Q7Z742;  
 DT 01-OCT-2003 (Tremblrel. 25, Created)  
 DT 01-OCT-2003 (Tremblrel. 25, Last sequence update)  
 DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)  
 DE Hypothetical protein.  
 OS Homo sapiens (Human)  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Testis;  
 RX MEDLINE=22388257; PubMed=12477932;  
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
 RA Brownstein M.J., Udwin T.B., Toshiyuki S., Carninci P., Prange C.,  
 RA Raha S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,  
 RA Bosak S.A., McSwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,  
 RA Whitting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,  
 RA Krzywinski M.I., Skalak U., Schmutz J., Schnerch A., Schein J.E.,  
 RA Jones S.J., Marra M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length human  
 and mouse cDNA sequences";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16999-16903 (2002).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Testis;  
 RA Strausberg R.;  
 RL Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; BC052966; AAH52966.1; -;  
 KW Hypothetical protein.  
 SQ SEQUENCE 976 AA; 47C8EE1826C4A74A CRC64;  
 Query Match 90.58; Score 4689.5; DB 4; Length 976;  
 Best Local Similarity 90.44; Pred. No. 0;  
 Matches 882; Conservative 29; Mismatches 64; Indels 1; Gaps 1;  
 QY 1 MSAGGCEPCTKPRWDAATSPPAASDARSFPQQRRLVDSKADPVQFRVPPSSGAL 60  
 DB 1 MNAGGCEPCTKPRWGAATTS-PAASDARSFPQQRRLVDSKADPVQFRVPPSSPACVP 59  
 QY 61 GRAGQHGSAATSLVFQKQITTSWMDTKIGTVESLSHKNNTRESMMSSVOKDNFY 120  
 DB 60 GRAGQHGSAATSLVFQKQITTSWMDTKIGTAESLSHKNNTRESMMSSVOKDNFY 119  
 QY 121 QHNMEKLENVQLGFDKSPVEKGTQYLKQHTAAWCKQWNGPHSELPPEAVTLVP 180  
 DB 120 QHNVEKLENVQLSLDKSPTEKSTQYLNQHTAAWCKQWNGKHTQLLESEPTVTLVP 179

QY 181 EOPSNANVDQSSPKDDHSDTNSERDNDQOFLTHVKLANAKOTMEDEOGEARSHQKCSK 240  
 DB 180 EOPSNANIDRSPQNDHSDTNSERDNDQOFLTHVKLANAKOTTEDEOGEARSHQKCSK 239  
 QY 241 ACHPAEACAGCQOEBTDDVVSPLSDTGSDEVTGTLKVNRLNRQESSLGNSSPPPEKES 300  
 DB 240 SCDPGECDACQOEBTDDVVSPLSDTGSDEVTGTLKVNRLNRQESSLGNSSPPPEKES 299  
 QY 301 PESPMVDVNSKNSCQDSADDETSPPGDEOEDSSAQTANKPSRQPPREADTELKRS 360  
 DB 300 PESPMVDVNSKNSCQDSADDETSPPGDEOEDSSAQTANKPSRQPPREADTELKRS 359  
 QY 361 KGGELRHFOPEGESRACMDVNAKRGSTSLNVECNKSKHGRKOSKIDHFMVVK 420  
 DB 360 KGGELRHFOPEGESRACMDVNAKRGSTSLNVECNKSKHGRKOSKIDHFMVVK 419  
 QY 421 AEDKKEOCEMKHQRTERKIPKYPHLSPPDKWLTGPIEEMRMPCGIRLPPRPSAN 480  
 DB 420 AEDKKEOCEMKHQRTERKIPKYPHLSPPDKWLTGPIEEMRMPCGIRLPPRPSAN 479  
 QY 481 HTVTIRVDLLRIGRVPKPPPTFKDILWKNHVMPCSEONLYPVEDENGERRAGSWELI 540  
 DB 480 HTVTIRVDLLRIGRVPKPPPTFKDILWKNHVMPCSEONLYPVEDENGERRAGSWELI 539  
 QY 541 QTALLNRLTRPQNLKDAILKYNVAYSKKWDFTALIDFWDKVLLEAEAOHLQSIILPDMVK 600  
 DB 540 QTALLNRLTRPQNLKDAILKYNVAYSKKWDFTALIDFWDKVLLEAEAOHLQSIILPDMVK 599  
 QY 601 IALCLPNICTQIPILLKQWNSHITMSQBIASLLANAFCTPPPRNAKMKSEYSSYPDI 660  
 DB 600 IALCLPNICTQIPILLKQWNSHITMSQBIASLLANAFCTPPPRNAKMKSEYSSYPDI 659  
 QY 661 NFNRLFEGSSRKPEKLTLCFVFRVTEKKTGLVTFTRQSLDFPWEKCEKLTLLRH 720  
 DB 660 NFNRLFEGSSRKPEKLTLCFVFRVTEKKTGLVTFTRQSLDFPWEKCEKLTLLRH 719  
 QY 721 VYVEGTIEGNGQMLQVDNFANFVGCVTSAGLVQBEIRFLINPELIVSRFLTEVLDHNE 780  
 DB 720 VYVEGTIEGNGQMLQVDNFANFVGCVTSAGLVQBEIRFLINPELIVSRFLTEVLDHNE 779  
 QY 781 CLITTEQYSEYTGATYRWARSHEDSRDMDWRRTTEIVAIDALHRRYLDQFVPE 840  
 DB 780 CLITTEQYSEYTGATYRWARSHEDSRDMDWRRTTEIVAIDALHRRYLDQFVPE 839  
 QY 841 KIRRELKAYCGFLRPGVSSSENLSAVATGNWCGAFGGDARLKALIQILAAVAERDVVY 900  
 DB 840 KMRRELKAYCGFLRPGVSSSENLSAVATGNWCGAFGGDARLKALIQILAAVAERDVVY 899  
 QY 901 FTFGSELMDIYSMTFLTERKLTGVEVYKLLRYNEECRCNSTPGDILKYPFIYHA 960  
 DB 900 FTFGSELMDIYSMTFLTERKLTGVEVYKLLRYNEECRCNSTPGDILKYPFIYHA 959  
 QY 961 VESCTQTTNPGQRTG 976  
 DB 960 VESCAETADHSQRTG 975  
 RESULT 4  
 Q9Y4W7 PRELIMINARY; PRT; 976 AA.  
 ID Q9Y4W7;  
 AC Q9Y4W7;  
 DT 01-NOV-1999 (Tremblrel. 12, Created)  
 DT 01-NOV-1999 (Tremblrel. 12, Last sequence update)  
 DT 01-JUN-2003 (Tremblrel. 24, Last annotation update)  
 DE Poly(ADP-ribose) glycohydrolase.  
 GN HPARG.  
 OS Homo sapiens (Human)  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.

Thu May 27 09:56:02 2004

```

RX MEDLINE=99380098; PubMed=10449915;
RA Ame J.C., Apou F., Jacobson E.L., Jacobson M.K.;
RT "Assignment of the poly(ADP-ribose) glycohydrolase gene (PARG) to
RT human chromosome 10q11.23 and mouse chromosome 14B by in situ
RT hybridization.";
RL Cytogenet. Cell Genet. 85:269-270(1999).
DR EMBL; AF005043; AAB61614.1; -.
DR Genbank; HGNC:8605; PARG.
DR GO; GO:0005737; C:cytoplasm; TAS.
DR GO; GO:0004649; F:poly(ADP-ribose) glycohydrolase activity; TAS.
DR InterPro; IPR007724; PARG.
DR Pfam; PF05028; PARG; 1.
KW Hydrolase.
SQ SEQUENCE 976 AA; 110903 MW; 42A0F51466FCE7DC CRC64;

Query Match 89.4%; Score 4635.5; DB 4; Length 976;
Best Local Similarity 89.5%; Pred. No. 0;
Matches 874; Conservative 32; Mismatches 69; Indels 1; Gaps 1;

QY 1 MSAGGCEPCTKRPWDAAATSPPAASDARSPPGRRVLDKAPVQFRVPPSSGCGAL 60
DB 1 MNAGGCEPCTKATGGAATTS-PAASDARSPPGRRVLDKAPVQFRVPPSPACVP 59

61 GRAGHRSATSLVFKOKTITTSWMDTKIGITVSESLHSKENNTRREESMMSSVQKDNFY 120
DB 60 GQAGHRSATSLVFKOKTITTSWMDTKIGITVSESLHSKENNTRREESMMSSVQKDNFY 119

121 QHNMEKLENVSLQGFDPKSPVEKGTQYLKQHTAAMCKWQNGPHERLLESEPPAVTLVP 180
DB 120 QHNVEKLVNSQLSLDKSLTEKSTQYLNOHTAAMCKWQNGEKTPEQLLESEPPAVTLVP 179

181 EQFSNANVDQSPKDDHSDTNSSESDNQOELTHVKLANAKOTMEDEGRRARSHQKCGK 240
DB 180 EQFSNANVDQSPKDDHSDTNSSESDNQOELTHVKLANAKOTMEDEGRRARSHQKCGK 239

241 ACHPAEACAGCQOETDVVSEPLSDTGSDEVTGTLKNNANLNROESSLGNSSPPPEKSE 300
DB 240 SCHPEDCASCQOEDIDVVPKSPLDVSGSDVTGSKNDKLIQESCLGNSSPPPEKSE 299

301 PESPMVDVNSKNSQDSEADETSPGDFQEDSSSAQTANKSRPQPRADTELKRSAA 360
DB 300 PESPMVDVNSKNSQDSEADETSPGDFQEDSSSAQTANKSRPQPRADTELKRSAA 359

361 KGEIRLHQFEGGESRAGMNDVNAKPGSTSSLVNVECRNSKHGRKDSKITDHFMRVPK 420
DB 360 KGEVRLHQPFGGESRTGMDNLAKLPGNISLVNVECRNSKHGRKDSKITDHLMLRPEK 419

421 AEDKRKEQCEMKHQRTERKIPKYPHLSPPDKKWLGTPIEEMRMPRCGIRLPLRPSAN 480
DB 420 AEDRRKEQWETKHQRTERKIPKYPHLSPPDKKWLGTPIEEMRMPRCGIRLPLRPSAN 479

481 HTVTIRVDLLRIGEVKPPPTTFKDIWDNKHVQMPCEQNLVPVEDENGERRAGSWELT 540
DB 480 HTVTIRVDLLRAGEVKKPPPTTHYKDLWDNKHVQMPCEQNLVPVEDENGERRAGSWELT 539

541 QTALLNRLRPQNLKDALIKYNVAYSKWDFALIDFWDKVLAEAAHQLYQSILPDMVK 600
DB 540 QTALLNKRFPQNLKDALIKYNVAYSKWDFALIDFWDKVLAEAAHQLYQSILPDMVK 599

601 IALCLPNICTQPIPLKQKNHSITWSQEQIASLLANAFCTTPRRNAKMSYSYPDI 660
DB 600 IALCLPNICTQPIPLKQKNHSITWSQEQIASLLANAFCTTPRRNAKMSYSYPDI 659

661 NFNRLPFGSRSRPEKLTUFCYFRRVTEKPTGLVTFTRQSLDEDPENWERCCKLTLH 720
DB 660 NFNRLPFGSRSRPEKLTUFCYFRRVTEKPTGLVTFTRQSLDEDPENWERCCKLTLH 719

721 VTYEGTTEGQOQMLQVDFANRFVGGGVTSAGLVQEEIRFLINPELLVSLFTVELDNE 780
DB 720 VTYEGTTEGQOQMLQVDFANRFVGGGVTSAGLVQEEIRFLINPELLVSLFTVELDNE 779

781 CLITITGEQSEYTGVAETRWRSRSHEDGSRDDCERRCTEIVADLHFRFVLDQFVPE 840

```

```

780 CLITITGEQSEYTGVAETRWRSRSHEDGSRDDCERRCTEIVADLHFRFVLDQFVPE 839
841 KIRRELKAYCGFLRPGVSSSENLSAVATGNWCGAFGGDARLKALIQILAAVAERDVVY 900
840 KMRELKAYCGFLRPGVSSSENLSAVATGNWCGAFGGDARLKALIQILAAVAERDVVY 899
901 FTFGDSELMRDIIYSMTFTFLTERKLTGVEVYKLLIRYNEECRCNSTGPDIKLYPFIYHA 960
900 FTFGDSELMRDIIYSMTFTFLTERKLTGVEVYKLLIRYNEECRCNSTGPDIKLYPFIYHA 959
961 VESCTQTTPGQORTG 976
960 VESCAETADSHGQRTG 975

RESULT 5
O88622 PRELIMINARY; PRT; 968 AA.
ID O88622;
AC O88622;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Poly(ADP-ribose) glycohydrolase.
GN PARG.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A. PubMed=10449915;
RX MEDLINE=99380098; PubMed=10449915;
RA Ame J.C., Apou F., Jacobson E.L., Jacobson M.K.;
RT "Assignment of the poly(ADP-ribose) glycohydrolase gene (PARG) to
RT human chromosome 10q11.23 and mouse chromosome 14B by in situ
RT hybridization.";
RL Cytogenet. Cell Genet. 85:269-270(1999).
DR EMBL; AF079557; AAC28735.1; -.
DR MGD; MGI:1347094; Parg.
DR GO; GO:0016787; F:hydrolase activity; IEA.
DR InterPro; IPR007724; PARG.
DR Pfam; PF05028; PARG; 1.
KW Hydrolase.
SQ SEQUENCE 968 AA; 109169 MW; 3981DA4809E1E007 CRC64;

Query Match 83.6%; Score 4332.5; DB 11; Length 968;
Best Local Similarity 84.0%; Pred. No. 1e-303;
Matches 820; Conservative 57; Mismatches 90; Indels 9; Gaps 5;

QY 1 MSAGGCEPCTKRPWDAAATSPPAASDARSPPGRRVLDKAPVQFRVPPSSGCGAL 60
DB 1 MSAGGCEPCTK-ARWGAAGTASPTASDSRSPGRRVLDKAPVQFRVPPSPACVS 59

61 GRAGHRSATSLVFKOKTITTSWMDTKIGITVSESLHSKENNTRREESMMSSVQKDNFY 120
DB 60 GRAGHRSATSLVFKOKTITTSWMDTKIGITVSESLHSKENNTRREESMMSSVQKDNFY 116

121 QHNMEKLENVSLQGFDPKSPVEKGTQYLKQHTAAMCKWQNGPHERLLESEPPAVTLVP 180
DB 121 QHNMEKLENVSLQGFDPKSPVEKGTQYLKQHTAAMCKWQNGPHERLLESEPPAVTLVP 176

117 PHKVEKLENVSLQGFDPKSPVEKGTQYLKQHTAAMCKWQNGPHERLLESEPPAVTLVP 176
DB 117 PHKVEKLENVSLQGFDPKSPVEKGTQYLKQHTAAMCKWQNGPHERLLESEPPAVTLVP 176

181 EQFSNANVDQSPKDDHSDTNSSESDNQOELTHVKLANAKOTMEDEGRRARSHQKCGK 240
DB 181 EQFSNANVDQSPKDDHSDTNSSESDNQOELTHVKLANAKOTMEDEGRRARSHQKCGK 233

177 KQLSNANIGQSPHTDDHSDTNSSESDNQOELTHVKLANAKOTMEDEGRRARSHQKCGK 233
DB 177 KQLSNANIGQSPHTDDHSDTNSSESDNQOELTHVKLANAKOTMEDEGRRARSHQKCGK 233

241 ACHPAEACAGCQOETDVVSEPLSDTGSDEVTGTLKNNANLNROESSLGNSSPPPEKSE 300
DB 241 ACHPAEACAGCQOETDVVSEPLSDTGSDEVTGTLKNNANLNROESSLGNSSPPPEKSE 293

301 PESPMVDVNSKNSQDSEADETSPGDFQEDSSSAQTANKSRPQPRADTELKRSAA 360
DB 301 PESPMVDVNSKNSQDSEADETSPGDFQEDSSSAQTANKSRPQPRADTELKRSAA 352

294 PESPMVDVNSKNSQDSEADETSPGDFQEDSSSAQTANKSRPQPRADTELKRSAA 352
DB 294 PESPMVDVNSKNSQDSEADETSPGDFQEDSSSAQTANKSRPQPRADTELKRSAA 352

361 KGEIRLHQFEGGESRAGMNDVNAKPGSTSSLVNVECRNSKHGRKDSKITDHFMRVPK 420

```



Db 353 KGSEVRLHFQFE - GENNACTSDLNAPKSGNSSLNVECRSSKHGKRDSKI THFMRISK 411  
QY 421 AEDKKEQCEMKHQRTERKIPKYPHLSDPDKWLGTPTEEMRMPCGIRLPPLRPSAN 480  
Db 412 SEDREKEQCEVHRQTERKIPKYPHLSDPDKWLGTPTEEMRMPCGIRLPPLRPSAS 471  
QY 481 HTVTIRVDLRLRAGEVPKPTTHFKDLDNKHVMPQSEQLNYPVEDENGERAGSRWELI 540  
Db 472 HTVTIRVDLRLRAGEVPKPTTHFKDLDNKHVMPQSEQLNYPVEDENGERAGSRWELI 531  
QY 541 QTALLNRLTRPQNLKDAIKYNVAYSKKWDFTALIDFWDKVLBEAAQHLIYQSILPDMVK 600  
Db 532 QTALLNKFTRPQNLKDAIKYNVAYSKKWDFTALIDFWDKVLBEAAQHLIYQSILPDMVK 591  
QY 601 IALCLPNICTQPIPLLKQNMHSITMSOEQIASLLANAFCTPRRNAMKMSYSSYPDI 660  
Db 592 IALCLPNICTQPIPLLKQNMHSITMSOEQIASLLANAFCTPRRNAMKMSYSSYPDI 651  
QY 661 NFNRLEPGRSSRKPEKLTLCYFRRVTEKKPTGLVTFTRQSLEDPEWERCCKPLTRLH 720  
Db 652 NFNRLEPGRSSRKPEKLTLCYFRRVTEKKPTGLVTFTRQSLEDPEWERCCKPLTRLH 711  
QY 721 VTVEGTIEGNGQMLQVDFANRFGVGGVTAGLVQBEIRFLINPELIVSRLEFTEVLDHNE 780  
Db 712 VTVEGTIEGNGQMLQVDFANRFGVGGVTAGLVQBEIRFLINPELIVSRLEFTEVLDHNE 771  
QY 781 CLIIITGEQYSEYTGAEYRWARSHEDSRDDQRRCTEIVAIIDALHFRYLDQFVPE 840  
Db 772 CLIIITGEQYSEYTGAEYRWARSHEDSRDDQRRCTEIVAIIDALHFRYLDQFVPE 831  
QY 841 KIRRELKAYCGFLRPGVSSENLSAVATGNWCGAFGCDARLKALIOILAAVAERDVVY 900  
Db 832 KVRRELKAYCGFLRPGVSSENLSAVATGNWCGAFGCDARLKALIOILAAVAERDVVY 891  
QY 901 FTFGDSSELMRDIYSMHTFLTERKLTVEGVYKLLRYNEECRCNSTPGPDIKLYPFIYHA 960  
Db 892 FTFGDSSELMRDIYSMHTFLTERKLTVEGVYKLLRYNEECRCNSTPGPDIKLYPFIYHA 951  
QY 961 VESCTQNTNPGORTG 976  
Db 952 VESSAETTDMPGQKAG 967  
RESULT 6  
Q80YQ6 PRELIMINARY; PRT; 961 AA.  
AC Q80YQ6;  
DT 01-JUN-2003 (TrEMBLrel. 24, Created)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Poly (ADP-ribose) glycohydrolase.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6; TISSUE=Brain;  
RA Strausberg R.;  
RL Submitted (APR-2003) to the EMBL/GenBank/DBJ databases.  
DR EMBL; BC050892; AAH50892.1; -  
DR GO; GO:0016787; F:hydrolase activity; IEA.  
DR InterPro; IPR007724; PARG.  
DR Pfam; PF05028; PARG; 1.  
KW Hydrolase.  
SQ SEQUENCE 961 AA; 108580 MW; 1558F120C367CF63 CRC64;  
Query Match 83.5%; Score 4329; DB 11; Length 961;  
Best Local Similarity 84.4%; Pred. No. 1.8e-303;  
Matches 818; Conservative 55; Mismatches 88; Indels 8; Gaps 4;  
QY 1 MSAGPGCEPCTKRPWDAAATSPPAASDARSFPGRRVLDSDKAPVQFRVPPSSGCA 60

Db 1 MSAGPGCEPCTKRPWDAAATSPPAASDARSFPGRRVLDSDKAPVQFRVPPSSPACVS 60  
QY 61 GRAGHRSATSLSVFKQKTIITSMMDTKGKTVYESSLSKNNNTREESMMSSVQKDNFY 120  
Db 61 GRAGPHRGNATSVFVKQKTIITSMMDTKGKTAESE---SKENNTRIDSMSSVQKDNFY 117  
QY 121 QINMEKLENVSGIFDKSPVEKGTQYLKQHOTAAAMCKWQNEGPHSERLLESBPATVLP 180  
Db 118 PHKVEKLENVPGINLIDKSPTEKSSQYLNQOQTASVCKWQNEGPHSERLLESBPATVLP 177  
QY 181 EFGSNANVDQSSPKDDHSDTNEESRDNOQFLTHVKLANAKQTMEDEQGEARSHQCGK 240  
Db 178 KQLSNANIGQSPTHDDSDTDEEDDNOQFLTPIKLANTKPTVGDQ---ARSNCKCSG 234  
QY 241 ACHPAACAGCQOEEDVVSSEPLSDTGSDEVTGKINANRLNRQESSLGNSSPPEKSE 300  
Db 235 SRQSVKDCGCGQEEVDVLPESPLSDVGAEDITGPKNDNKLGTQESSLGDSPPEKSE 294  
QY 301 PESPMVDNKSNCODEADEETSPGDFDQEDSSSAQTANKPSRFPQPREADTELKRSSA 360  
Db 295 PESPMVDNKSNCODEADEETSPVDEQDDRSS--QTANKLSSCOAREADGLRXYLT 353  
QY 361 KGGEIRLHFQFEGESRAGMNDVNAKPGSTSSINVECRNSKQHGKRDSKI THFMRVPK 420  
Db 354 KGSEVRLHFQFE - GENNVGTSDLNAPKSGNSSLNVECRSSKHGKRDSKI THFMRISK 412  
QY 421 AEDKKEQCEMKHQRTERKIPKYPHLSDPDKWLGTPTEEMRMPCGIRLPPLRPSAN 480  
Db 413 SEDRRKEQCEVHRQTERKIPKYPHLSDPDKWLGTPTEEMRMPCGIRLPPLRPSAS 472  
QY 481 HTVTIRVDLRLRAGEVPKPTTHFKDLDNKHVMPQSEQLNYPVEDENGERAGSRWELI 540  
Db 473 HTVTIRVDLRLRAGEVPKPTTHFKDLDNKHVMPQSEQLNYPVEDENGERAGSRWELI 532  
QY 541 QTALLNRLTRPQNLKDAIKYNVAYSKKWDFTALIDFWDKVLBEAAQHLIYQSILPDMVK 600  
Db 533 QTALLNKFTRPQNLKDAIKYNVAYSKKWDFTALIDFWDKVLBEAAQHLIYQSILPDMVK 592  
QY 601 IALCLPNICTQPIPLLKQNMHSITMSOEQIASLLANAFCTPRRNAMKMSYSSYPDI 660  
Db 593 IALCLPNICTQPIPLLKQNMHSITMSOEQIASLLANAFCTPRRNAMKMSYSSYPDI 652  
QY 661 NFNRLEPGRSSRKPEKLTLCYFRRVTEKKPTGLVTFTRQSLEDPEWERCCKPLTRLH 720  
Db 653 NFNRLEPGRSSRKPEKLTLCYFRRVTEKKPTGLVTFTRQSLEDPEWERCCKPLTRLH 712  
QY 721 VTVEGTIEGNGQMLQVDFANRFGVGGVTAGLVQBEIRFLINPELIVSRLEFTEVLDHNE 780  
Db 713 VTVEGTIEGNGQMLQVDFANRFGVGGVTAGLVQBEIRFLINPELIVSRLEFTEVLDHNE 772  
QY 781 CLIIITGEQYSEYTGAEYRWARSHEDSRDDQRRCTEIVAIIDALHFRYLDQFVPE 840  
Db 773 CLIIITGEQYSEYTGAEYRWARSHEDSRDDQRRCTEIVAIIDALHFRYLDQFVPE 832  
QY 841 KIRRELKAYCGFLRPGVSSENLSAVATGNWCGAFGCDARLKALIOILAAVAERDVVY 900  
Db 833 KVRRELKAYCGFLRPGVSSENLSAVATGNWCGAFGCDARLKALIOILAAVAERDVVY 892  
QY 901 FTFGDSSELMRDIYSMHTFLTERKLTVEGVYKLLRYNEECRCNSTPGPDIKLYPFIYHA 960  
Db 893 FTFGDSSELMRDIYSMHTFLTERKLTVEGVYKLLRYNEECRCNSTPGPDIKLYPFIYHA 952  
QY 961 VESCTQNTN 969  
Db 953 VESSAETTD 961  
RESULT 7  
Q80YQ6 PRELIMINARY; PRT; 972 AA.  
ID Q80YQ6  
AC Q80YQ6;  
DT 01-MAY-2000 (TrEMBLrel. 13, Created)

DT	01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT	01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE	Poly(ADP-ribose) glycohydrolase.
GN	PARG.
OS	Rattus norvegicus (Rat).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX	NCBI_TaxID=10116;
OX	[1]
RN	SEQUENCE FROM N.A.
RC	STRAIN=BUF; TISSUE=Colon;
RA	Shimokawa T., Masutani M., Nagasawa S., Nozaki T., Ikota N., Araki S.,
RA	Aoki Y., Nakagama H., Sugimura T.;
RT	"Isolation and cloning of rat poly(ADP-ribose) glycohydrolase
RL	(Parg)."; (OCT-1998) to the EMBL/GenBank/DBJ databases.
DR	EMBL; AB019366; BAA87901.1; -.
DR	GO; GO:0016787; F:hydrolyase activity; IEA.
DR	InterPro; IPR007724; PARG.
DR	Pfam; PF05028; PARG; 1.
KW	Hydrolase.
SQ	SEQUENCE 972 AA; 109002 MW; 3DED7885684F59E9 CRC64;
Query Match	82.9%; Score 4295.5; DB 11; Length 972;
Best Local Similarity	82.6%; Pred. No. 4.9e-301;
Matches	808; Conservative
QY	1 MSAGPGCEPTKRRPWDAATSPPAASDARSFPQRORVLDSKDAPVQFRVPSSSGCAL 60
DB	1 MSAGPGCEPTKRRPWDAATSPPAASDARSFPQRORVLDPKADPVQFRVPSSSACVS 60
QY	61 GRAGHGRSATSLVFVKOKTITSMWDTKGIKTVESSELHSKNNNTREESMMSSVQKNFY 120
DB	61 GRAGPHRGSVTSFVKOKPIITWMDTKGPKTAESE--SKENNTRTDPMMSVQXDNFY 117
QY	121 QHNMEKLNVSQLGDFDKSPVEKGTQYLKHOTAAWKQWNGEPHSERLLESEPVALTLVP 180
DB	118 PHKVEKLVNPQLNLDSKFTKSTPYLNQQQTAGVCKVKSAGERAEQSASEPVAVTQP 177
QY	181 EQFSNANVDQSPKDDHSDTNSSESRDNOQFLTHVKLANAKQTWEDEQREARSHQCGK 240
DB	178 KQLSNANIQSPPTDGHSDTTHEEDRDNQFLTPVKLANAKQTVDGQ---ARSNCKCSA 234
QY	241 ACHPAEACAGCQOEETDVVSESPLDTSGETVGTGLKVNRINROESSLGNSPPPEKESE 300
DB	235 SCQCQDCAGCCQREADVIPESPLSDVAEDIGTGSKNDNKLTGQESGLGSDPPPEKESE 294
QY	301 PESPMVDVNSKSCQSEADETSFGDFQDESS--SAQTANKPSRFOPREADTELKRYS 358
DB	295 PESPMVDVNSKTSQSEADEASPFVFDQDDQDDRSSQTANKLSRQAQEVGDLRKKY 354
QY	359 SAKGGIRLHFQEGGESBAGMVDNAKRPGSTSLNVCRNSKQHKRKDSKITDHPFRV 418
DB	355 LTKGEIRLHFQEGG-SNAGTSDLNAKPSGNSSLNVDGRSSKQHGKDSKITDHPVRI 413
QY	419 PXAEDRRKBQCEMKHQRTKPKYIYPHLSPDKWLGPTEEMRMWRPCGIRLPPLRPS 478
DB	414 PKSEDRRKQEVRHQRAERKPKYIYPPLPPDKWLGPTEEMRMWRPCGIRLPPLRPS 473
QY	479 ANHTVTIRVDLLRIGVPEKPPPTHFKDLMDNKHVMKPCSEONLYPEDENGERRAGSRWE 538
DB	474 ASHTVTIVRVDLLRAGEVPEKPPPTHYKDLMDNKHVMKPCSEONLYPEDENGERTAGSRWE 533
QY	539 LIQTALNRLTRPNLKDALIKYNVASKWDFTALIDFWDKVLEBAEAQHLYQSLTLPDM 598
DB	534 LIQTALINKLTRPNLKDALIKYNVASKWDFETALVDFDKVLEBAEAQHLYQSLTLPDM 593
QY	599 VKIALCLPNTCTQPIPLLKOMHISITWSQEQIASILANAFFCTFFPRNAKMSEYSSYP 658
DB	594 VKIALCLPNTCTQPIPLLKOMHISVITWSQEQIASILANAFFCTFFPRNAKMSEYSSYP 653
QY	659 DINFNRLFEGRSRRKPEKLTLCFYFRFRVTEKKPTGLVTFTRQSLIEDFPPEWERCKPLTR 718
DB	

```

Db      235  SRQSVKCTGCGQEEVVDVLPSPSLSDVCAEDIGTGPKNKDLTGQESSIGDSPPEKSE 294
QY      301  PESPMVDVNSKNSQDSSEADETSPGFDEQEDSSAQTANKPSRFPQPREATELKRSSA 360
Db      295  PESPMVDVNSKNSQDSSEADETSPGFDEQEDSSAQTANKLSSQAREADGDLKRYLT 353
QY      361  KGGEIRLHFQEGESRAGMDVNAKRPGSTSSLVNVECRNSQKQKSKITDHFMRVPK 420
Db      354  KGSEVRLHFOFE-GENNAGTSLNAKPSGSSLVNVECRNSQKQKSKITDHFMRISK 412
QY      421  AEDKEKECEMKHQHTEKIPKYPHPSLSPDKKWLGTPIEENRRMPRGIRLPLRPSAN 480
Db      413  SEDRRKQCEVHQHTEKIPKYPHPSLSPDKKWLGTPIEENRRMPRGIRLPLRPSAS 472
QY      481  HTVTIRVDLLRIGEVKPPFTHFKOLMDNKHVMPCSEONLYPVDENGERAGRWELI 540
Db      473  HTVTIRVDLLRIGEVKPPFTHFKOLMDNKHVMPCSEONLYPVDENGERTAGRWELI 532
QY      541  QTALLNRLTRPQNLKDAILKYNVASKWDFDALIDFWDKVLLEAEAOHLYSILPDMVK 600
Db      533  QTALLNRLTRPQNLKDAILKYNVASKWDFDALIDFWDKVLLEAEAOHLYSILPDMVK 592
QY      601  IALCLPNICTQIPILKQKNHSHITMSQEQIASLLANAFCTFPRRNAMKSEYSYPDI 660
Db      593  IALCLPNICTQIPILKQKNHSHITMSQEQIASLLANAFCTFPRRNAMKSEYSYPDI 652
QY      661  NFNRLEFEGSSRKPEKLTLCYFRRVTEKPTGLVTFQSLDEDPWECEKLLTRLH 720
Db      653  NFNRLEFEGSSRKPEKLTLCYFRRVTEKPTGLVTFQSLDEDPWECEKPLTRLH 712
QY      721  VTYEGTIEGNGOMLOVDPANFVGSGVTSAGLVQSEIRPLNPELIVSRKLTVEVLHNE 780
Db      713  VTYEGTIEGNGOMLOVDPANFVGSGVTSAGLVQSEIRPLNPELIVSRKLTVEVLHNE 772
QY      781  CLIIITGEQVSEYTGVAETVWRASHEDRSERDDWQRRTEIIVADLHFRYLDQFVPE 840
Db      773  CLIIITGEQVSEYTGVAETVWRASHEDRSERDDWQRRTEIIVADLHFRYLDQFVPE 832
QY      841  KIRRELKAYCGFLRPGVSENLSAVATGNWGCAGFGGDARLKALIQILAAAVERDVVY 900
Db      833  KVRRELKAYCGFLRPGVSENLSAVATGNWGCAGFGGDVRLKALIQILAAAVERDVVY 892
QY      901  FTFGSELMDIYSMTHTLTKLTWGE 928
Db      893  FTFGSELMDIYSMTHTLTKLTWGE 920

RESULT 9
O46043  PRELIMINARY; PRT; 768 AA.
AC      O46043;
DT      01-JUN-1998 (TrEMBLrel. 06, Created)
DT      01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT      01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE      PARG protein.
GN      PARG OR EG:114E2.1 OR CG2864.
OS      Drosophila melanogaster (fruit fly).
OC      Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC      Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC      Ephydroidea; Drosophilidae; Drosophila.
OX      NCBI_TaxID=7227;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=Berkeley;
RX      MEDLINE=20196006; PubMed=10731132;
RA      Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA      Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA      George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA      Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA      Brandon R.C., Rogers Y.-H.C., Blazek R.G., Champe M., Pfeiffer B.D.,
RA      Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA      Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA      Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,

```

```

RA      Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA      Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA      Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA      Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA      de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA      Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA      Durbin K.J., Evangelista C.C., Ferraz C., Ferreria S., Fleischmann W.,
RA      Foaier A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris K.,
RA      Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA      Hoscin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA      Jafali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA      Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA      Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA      Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA      Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA      Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA      Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA      Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA      Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA      Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA      Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA      Svirskaas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA      Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissenbach J.,
RA      Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA      Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA      Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA      Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT      "The genome sequence of Drosophila melanogaster.";
RL      Science 287:2185-2195 (2000).
RN      [2]
RP      SEQUENCE FROM N.A.
RA      Ame J.-C., Jacobson M.K.;
RT      "Isolation and characterization of the cDNA encoding Drosophila
RT      poly(ADP-ribose) glycohydrolase.";
RL      Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
RN      [3]
RP      SEQUENCE OF 46-768 FROM N.A.
RA      Murphy L., Harris D., Bazzell B.;
RL      Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.
RN      [4]
RP      SEQUENCE OF 46-768 FROM N.A.
RA      Benos P.;
RL      Submitted (DEC-1997) to the EMBL/GenBank/DBJ databases.
DR      EMBL; AF003428; AAC45886.1; -.
DR      EMBL; AF079556; AAC28734.1; -.
DR      EMBL; Z98254; CAB10913.1; -.
DR      FlyBase; FBgn0023216; Parg.
DR      GO; GO:0016787; F:hydrolase activity; IEA.
DR      InterPro; IPR007724; PARG.
DR      Pfam; PF05028; PARG; 1.
KW      Hydrolase.
SQ      SEQUENCE 768 AA; 86288 MW; 2E620C42E643FC04 CRC64;

Query Match      20.5%; Score 1063.5; DB 5; Length 768;
Best Local Similarity 41.5%; Pred. No. 6e-68;
Matches 244; Conservative 93; Mismatches 198; Indels 53; Gaps 15;

QY      417  RVPAEDKRKEQCEMKHQHTEKIPKYPHPSLSPDKKWLGTPIEENRRMPRGIRL 472
Db      45  RMSKSPDGGGISEIETEE-----PENLANSLLDSNRGVSMETHNRQPFLENL 94
QY      473  PPLRPSANHTVTIRVDLLRIGEVKPPFTHFKOLMDNKHVMPCSEONLYPVDENGERA 532
Db      95  PPVTAGNLHRVMYQ---LPIRETP-PRPYKSPGKWDSEHVRFLPCAPESKYPRENPDG 150
QY      533  AGRWELIQTALLNRLTRPQNLKDAILKYNVASKWDFDALIDFWDKVLLEAEAOHL 592
Db      151  IDFRWEMIERALLQPIKTCBELQAIIISYNTTRDQWHFRALHQLLDELDSETRVFE 210
QY      593  SILPDMVKIALCLPNICTQIPILKQKNHSHITMSQEQIASLLANAFCTFPRRNAM 651
Db      211  DLLPRIIRLALRDLPLTQSPVPLKHKHNASLSLSQOISCLLANAFCTFPRRNAM 270

```



```

Db      139 VFLSQELIGALLACGFFCLFPDDNRGAK-----HLPVINFDHLFASISYISOSQSKIRC 194
QY      680 LFCYFRVTEKKTGLVTTTQ--SLEDFFPWERCEKILTRHVTVEGTIEGNGQGMLOV 737
Db      195 IMHYFERFSCVPIGIVSFERKITAAPDADFWSKSDVSLCAPKHSFGLIEOPDPALEV 254
QY      738 DFANRFGGVTGAGVQBEIRFLINPELIVSRFTTEVDHNECLITTEQYSEYTGVA 797
Db      255 DFANKYLGGSLSRGCVQBEIRFMINPELITAGMLFLPRMDNEAIEIVGAERFSCVTGYA 314
QY      798 EYRWARSHEDSRDWRRTTEIVAIDAL-----HFRYLDQVPEKIRRELNKAYC 851
Db      315 SGRFAGYIDKAMDFFKRRTRIVAIIDALCTPKMRHFKDIC-----LUREINKALC 367
QY      852 GFLR-----PGVSS-----ENLSAVA 867
Db      368 GFLNCSKANEHQNI FMEGDNEIQLVRNGRDSGLLTETTTASHRTPLNDVEMNREKPA 427
QY      856 -----PGVSSNL--SAVATNGWCGAGFGGDARUKALIQILAAVAERDVV-YFTFGD 905
Db      428 LTRDFYVEGDNEHDDGVATNGWCGVFGGDEPKATOWLAASQTRRPFISYTFG--486
QY      906 SELMRDIYSMHTFELTERKLTGGEVYKLLRY 936
Db      487 VEALRNLQVTKWLSHKWTGDLNMNMLEY 517

RESULT 12
Q8VYAL ID Q8VYAL PRELIMINARY; PRT; 522 AA.
AC Q8VYAL;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Putative poly(ADP-ribose) glycohydrolase.
DE AT2G31870.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Nguyen M., Karlin-Neumann G., Southwick A., Lam B., Miranda M.,
RA Palm C.J., Bowser L., Jones T., Banh J., Carninci P., Chen H.,
RA Cheuk R., Chung M.K., Hayashizaki Y., Ishida J., Kamiya A., Kawai J.,
RA Kim C., Lin J., Liu S.X., Narusaka M., Pham P.K., Sakano H.,
RA Sakurai T., Satou M., Seki M., Shinn P., Yamada K., Shinozaki K.,
RA Ecker J., Theologis A., Davis R.W.;
RA Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
RL EMBL; AY072330; AAL61937.1;
DR GO; GO:0016787; F:hydrolase activity; IEA.
DR InterPro; IPR007724; PARG.
DR Pfam; PF05028; PARG; 1.
KW Hydrolase.
SQ SEQUENCE 522 AA; 58942 MW; 6FF98685D867CA15 CRC64;

Query Match 10.3%; Score 533; DB 10; Length 522;
Best Local Similarity 30.5%; Pred. No. 6.6e-30;
Matches 132; Conservative 83; Mismatches 138; Indels 80; Gaps 11;

QY 577 FWDKVLBEAQAHHYQSILPMVKIALCLPNICTQPIPLKQKNH-----622
Db 80 FFDKKIKRESANFGEVWPALCRLLQLPSMLEKH-----QKADHVLGDGVKSLRLG 135
QY 623 -----SITWQEQIASILLANAFCTFPRNAOKSESYSDINENLFE----GRSRKP 674
Db 136 QEAGIVLLSQELIAALACSFCLFPVEVDRSLK-----NLQGINFSGLFSFPYMHCKTQE 191
QY 675 EKLATLFCYFRVTEKKTGLVTTTQSL-----DPFE---WERCEKILTRHVTY 723
Db 192 NKIKLHYFGRCRWMPGTGVSEFKILPLEYHPHFVSPKADSWANSVTPLCSIHT 251

```

```

QY 724 EGTIEGNGQMLQVDFANRFGVGGVTGAGVQBEIRFLINPELIVSRFTTEVDHNECL 783
Db 252 SGAIDQPCAELEVDFADEYFGGLTSLVDTLQBEIRFVINPELITAGMIFLPRMDANEAT 311
QY 784 IGTQYSEYTGAYETRWARSHEDSRDWRRTTEIVAIDALHFRYLDQVPEKIR 843
Db 312 IVGVERFSGYTGYSFGYAGDYTNKOLDIFRRKTRVIAIDAMP-DPGMGQYKLDALI 370
QY 844 RELNKAYCGFLR-----PGVSS-----ENLSAVA 867
Db 371 REVNKAFSGYMHQCKYNIDVHKDPEASSSHVPLTSDSASQVIESSHRWICDEEEKIGVA 430
QY 868 TGNWCGAGFGGDARUKALIQILAAVAERDVV-YFTFGDSELMRDIYSMHTFELTERKLT 926
Db 431 TGNWCGVGGDEPKATOWLAISQSGRPFMSYTFG--LQALQNLNQVEMVALQEMTV 489
QY 927 GEVYKLLRYNE 939
Db 490 GDLMKKLVESSE 502

RESULT 13
Q867X0 ID Q867X0 PRELIMINARY; PRT; 781 AA.
AC Q867X0;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Poly ADP-ribose metabolism enzyme-3 long form (C. elegans PME-3
DE protein) (Corresponding sequence F20C5.1a).
DE F20C5.1 OR PME-3.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Gagnon S.N., Hardy I., Desnoyers S.;
RA "Characterization of poly(ADP-ribose) glycohydrolases in the nematode
RA Caenorhabditis elegans.";
RL Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=99069613; PubMed=9851916;
RA none;
RT "Genome sequence of the nematode C.elegans: A platform for
RT investigating biology.";
RL Science 282:2012-2018(1998).
RN [3]
RP SEQUENCE FROM N.A.
RA Matthews P.;
RA Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.
RL EMBL; AY185493; AAO26316.1; -.
DR EMBL; Z68161; CAD89735.1; -.
DR WormBep; F20C5.1a; CE333775.
DR InterPro; IPR007724; PARG.
DR Pfam; PF05028; PARG; 1.
SQ SEQUENCE 781 AA; 89273 MW; 30447D06605CED0A CRC64;

Query Match 9.3%; Score 481.5; DB 5; Length 781;
Best Local Similarity 24.1%; Pred. No. 6.2e-26;
Matches 210; Conservative 126; Mismatches 297; Indels 239; Gaps 37;

QY 199 DTNSERDNRQOPLTHVKLANAKQTM-----DEQGRASHQKCGKACHPAEACAGQOE 254
Db 10 DPVTQDEKQYEDV-VGVFAHQVPTMKRRLTEHGNNTESKE-----DPEE-----PK 56
QY 255 ETDVSESPLSDTGSDE-----VGTGLKNA-----NRLNQESSLGN 291
Db 57 SRDVFVSSQSSDSQBSQBSAENPIAEKVENSCENTLTCLKISLNDNTERSEHTLDN 116
QY 292 ---SPFEKESEPESPMDVDNKNKSCQDSEA-----DEETSPGDEQED-----332

```

```

Db 117 HKSTEPMEEDVNNKSNIDV--AINSDDEDDLVLEENKEMKRDGEVQOQDLFADDQELLEY 174
Qy 333 -----SSSAQTANKPSRQPPREADTELKRSSAKGGEIRLHFQFEGGESRA 378
Db 175 PGIMKDTTQDITDSEVETAQKMEIEETEADS-----TFVGEDSK- 216
Qy 379 GMNDVNAKRPGST-----SSLNVECRNSKQKORHOKSDKITDHFMRVPKAEDEKKEQCEM 431
Db 217 -----NQSQSGTTSDEVDADSQINLATKTVR-----TSSSSFLSTVSTCEAPAKGRARM 265
Qy 432 KHQTERKIPKIYPHLS--PDKKWLGTPIEEMRMRPCGIRLPLPSPSANHTVIRVDL 489
Db 266 YQKELEKHVIAFTGNTLQPD-----LNKVPDPDRNY----- 297
Qy 490 LRIGEVKPPFP--THFKOLWDNKH--VKMPCSEONLYPVEDENGERRAAGSRWELIQTALL 545
Db 298 -RYCTIPN-FPASQGLREDNRYGPKIVLPQWRREF-----DSRGR- 337
Qy 546 NRLTRPQNLDAILKYN--VAYSQKWDFTA-----LIDFWD--KVLBEAQAHLHYQ 592
Db 338 -----RDSYFYFKKLDGKYLKCYKTYGYFMFVGLLHNMWEPDPDITYKLPALEMYI 388
Qy 593 SILPMV-----KIALCLPNICTOPIPLKQKXNHSITMSQEOIASILANAF 639
Db 389 KEMSELVGREVLEKPARVARIATAEDILPERIYRLVGDV-ESATLSHKQCAALVARMF 447
Qy 640 FCTTPRRNAKMSSESYPDINFRNLFGRSSRKPEKLTLCFYPRRVTEKKTGLVTF- 698
Db 448 F-----ARPDSPFS-----FCRILSSDSKICVEKLLKFTFYFDKMSMDPPDGAVSFR 494
Qy 699 -TROSLDPPWERCEKLLTRIHYV--EGTIEGNGQMLQVDFANRVFGGVTISAGLVO 755
Db 495 LTKMDKDTFNEEWKDKLRLSPEVEFFDEMLIEDTAL-CTQVDFANEHLGGVGLNHGSVQ 553
Qy 756 BEIRLIPNELIVSLFTEVLDHNECLITGTEOYSEYTVGAETVRWA-----RSHEDRS 810
Db 554 BEIRFLMCPMMVMGMLCEKMKQLEAISIVGAYVFSSTGYGHTLKNLQELQPNHRSQNTN 613
Qy 811 E-RDDQWRRTTRIVADALHFR------YLDQFVPEKIRRELKAYCYGFLRPGVSEMLS 864
Db 614 EFRDFRGLRVETIADAILFKGSKLDCTEQELKNANIIRKMKASIGFMSQGPFTNI- 672
Qy 865 AVATGNWGGAGGAGDARLQALQILAAVAERDVVYFFGDSLEMRDIYSMTFLTERKL 924
Db 673 PIVTGWGGAGFNGKPLKFIQVIAAGVADRLPHFCSPGEBELAACKKIIRMKQKDV 732
Qy 925 TVGEVYKLLRLRYNEECRNCSTPGDPKLYPF 956
Db 733 TLGMLFSMI-----NNTGLPHKHFEFYVP 756

```

RESULT 14

```

Q19637 PRELIMINARY; PRT; 764 AA.
AC Q19637;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE PME-3 protein (corresponding sequence F20C5.1) (Poly ADP-ribose
DE metabolism enzyme-3 short form).
GN F20C5.1 OR PME-3.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Matthews P.;
RL Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=99069613; PubMed=9851916;

```

```

RA none;
RT "Genome sequence of the nematode C.elegans: A platform for
RT investigating biology.";
RL Science 282:2012-2018(1998).
RN [3]
RP SEQUENCE FROM N.A.
RA Gagnon S.N., Hardy I., Desnoyers S.;
RT "Characterization of poly (ADP-ribose) glycohydrolases in the nematode
RT Caenorhabditis elegans.";
RL Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; Z68161; CA92299.2; --
DR EMBL; AY185494; AAO26317.1; --
DR PIR; T21138; T21138.
DR WormPep; F20C5.1b; CE32867.
DR InterPro; IPR007724; PARG.
DR Pfam; PF05038; PARG; 1.
SQ SEQUENCE 764 AA; 87441 MW; B391703B1D0CB79A CRC64;

Query Match 9.2%; Score 475; DB 5; Length 764;
Best Local Similarity 24.0%; Pred. No. 1.8e-25;
Matches 208; Conservative 120; Mismatches 294; Indels 244; Gaps 36;

Qy 199 DTNSESRDNOQFLTHYKLANAKTME-----DEQREARSHOKGKACHPAEACAGCQOE 254
Db 10 DPVTQDSEKDYEVV-GVGEFAHQVPTMKRKLTERGNTESKE-----DPEE-----PK 56
Qy 255 ETDVVSSEPLSDTGSED-----VGTGLKNA-----NRLNROESSIGN 291
Db 57 SRDVFVSSQSDSEQDSAPNPEAKVESENCENTLTKISNIESLDNVTERSEHTLDN 116
Qy 292 ---SPPEKSESESPMDV-----DNKNSQDSEADEETSPPF--DEOE--- 331
Db 117 HKSTEPMEEDVNNKSNIDVAINSDDEDDLVLEENKEMKRDGEVQOQDLFADDQELLEY 176
Qy 332 -----DSSAQATANKPSRQPPREADTELKRSSANGGEIRLHFQFEGGESR 377
Db 177 YPGIMKDTTQDITDSEVETAQKMEIEETEADS-----TFVGEDSK 219
Qy 378 AGMNDVNAKRPGSTSSLNVECRNSKQKORHOKSDKITDHFMRVPKAEDEKKEQCEMKHQTE 437
Db 220 A-----TKTVRSSSSFLST-----VSTCEAPAKGRARMYQKELE 254
Qy 438 RKIPKIYPHLS--PDKKWLGTPIEEMRMRPCGIRLPLPSPSANHTVIRVDLRLRIGEV 495
Db 255 KHVIAFTGNTLQPD-----LNKVPDPDRNY-----RYCTI 285
Qy 496 KPFPF-THFKOLWDNKH--VKMPCSEONLYPVEDENGERRAAGSRWELIQTALLNRLTRP 551
Db 286 PN-FPASQGLREDNRYGPKIVLPQWRREF-----DSRGR- 320
Qy 552 QMLKDAILKYN--VAYSQKWDFTA-----LIDFWD--KVLBEAQAHLHYQSLPDM 598
Db 321 ---RDSYFYFKKLDGKYLKCYKTYGYFMFVGLLHNMWEPDPDITYKLPALEMYKEMSEL 377
Qy 599 V-----KIALCLPNICTOPIPLKQKXNHSITMSQEOIASILANAFCTEPR 645
Db 378 VGREEVLEKPARVARIATAEDILPERIYRLVGDV-ESATLSHKQCAALVARMF----- 431
Qy 646 RNAKMSSESYPDINFRNLFGRSSRKPEKLTLCFYPRRVTEKKTGLVTF--TROSL 703
Db 432 ---ARPDSPFS-----FCRILSSDSKICVEKLLKFTFYFDKMSMDPPDGAVSFRLTQMDK 483
Qy 704 EDFPEWERCEKLLTRIHYV--EGTIEGNGQMLQVDFANRVFGGVTISAGLVOEIRFL 761
Db 484 DTFNEEWKDKLRLSPEVEFFDEMLIEDTAL-CTQVDFANEHLGGVGLNHGSVQOEIRFL 542
Qy 762 INPELIVSLFTEVLDHNECLITGTEOYSEYTVGAETVRWA-----RSHEDRS--RDDW 815
Db 543 MCPMMVMGMLCEKMKQLEAISIVGAYVFSSTGYGHTLKNLQELQPNHRSQNTNFRDRF 602
Qy 816 QRRTTEIIVADALHFR-----YLDQFVPEKIRRELKAYCYGFLRPGVSESLAVATGN 870
Db 603 GRLRVETIADAILFKGSKLDCTEQELKNANIIRKMKASIGFMSQGPFTNI-PIVTCW 661

```

689 EKPTGLVTF-----RQSLDFPEWE-RCEKLLRLHVTVEGTIEGNGQGLQVDFAN 741  
201 IDPPIGAVSFRKMRITHKOYLEN---WKLRETNLLPDVQVDFKMSIETAL-CTQIDFAN 256  
742 RFVGGVTSAGLVOEIRFLINPELIVSRFLTEVLDHNECLITTEQYSEYTGTAETVR 801  
257 KRLGGVLKGGAVQEEIRFMWCPENMVAILLNDVTQDLEAISIVGAYVFPSSYTGYSNTLK 316  
802 WAR-----SHEDRSERDDWOPRTTEIIVAIDALH-----FRRLDQFVPEKIRRELKAY 850  
317 WAKITPKHSAQNNSFRDQFGLQETETVAIDAVRNAGTPLECLLNQLTTEKLTREVKKAA 376  
851 CGFLRPGVSSSENLSAVATGNWCGGAFGGDARLKALIQILAAVAERDVVYFTFGDSELMR 910  
377 IGFLSAGDGFSKI-PVSGWNGCGAFRGNKPLKFLIQVIACGISDRPLQCFCTFGDTLAK 435  
911 DIYSMHTFLTERKLTVEVYKLLRYNEECNCSCTPGDPIKL-YFFIYHVESCTQTT 968  
436 KCEEMTLFRNNNRVTGQLFLII-----NSIGPPLNYSQYVDFDAIRAKINST 483

Search completed: May 26, 2004, 18:45:55  
Job time : 100.971 secs

QY 871 WCGGAFGGDARLKALIQILAAVAERDVVYFTFGDSELMRDIYSMHTFLTERKLTVEVY 930  
Db 662 WCGGAFNGDKPLKFLIQVIAGVADRLPHFCFSGPELAACKKIIERMKQDVLGLMLF 721  
QY 931 KLLRYNEECNCSCTPGDPIKLKYPF 956  
Db 722 SMI-----NNTGLPHKHFEFYVF 739

RESULT 15  
Q9NSL4 PRELIMINARY; PRT; 485 AA.  
AC Q9NSL4; AC Q9NSL4; 15, Created)  
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)  
DT 01-MAR-2003 (Tremblrel. 23, Last sequence update)  
DT 01-JUN-2003 (Tremblrel. 24, Last annotation update)  
DE Hypothetical protein H23L24.5 (Poly ADP-ribose metabolism enzyme  
DE 4).

GN H23L24.5 OR PME-4.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
OC Rhabditidae; Peloderinae; Caenorhabditis.  
OX NCBI\_TaxID=6239;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Bristol N2;  
RX MEDLINE=99069613; PubMed=9851915;

RA Waterston R.;  
RT "Genome sequence of the nematode C. elegans: a platform for  
RT investigating biology. The C. elegans Sequencing Consortium.";  
RL Science 282:2012-2018 (1998).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Bristol N2;  
RA Abu-Threideh J., Lehnert L.;  
RT "The sequence of C. elegans cosmid H23L24.";  
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Bristol N2;  
RA Waterston R.;  
RL Submitted (DEC-2002) to the EMBL/GenBank/DBJ databases.  
RN [4]

RP SEQUENCE FROM N.A.  
RA Gagnon S.N., Hardy I., Desnoyers S.;  
RT "Molecular characterization of poly ADP-ribose glycohydrolases in  
RT Caenorhabditis elegans.";  
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AC006662; AAF39896.2; -  
DR EMBL; AF548468; AAN40699.1; -  
DR WormPep; H23L24.5; CE32685.  
DR InterPro; IPR007724; PARG.  
DR Pfam; PF05028; PARG; 1.  
KW Hypothetical protein.  
SQ SEQUENCE 485 AA; 55574 MW; 86367DDBAB15081F CRC64;

Query Match 9.1%; Score 472; DB 5; Length 485;  
Best Local Similarity 29.9%; Pred No 1.5e-25;  
Matches 143; Conservative 83; Mismatches 165; Indels 88; Gaps 17;  
QY 537 WELTQALLNRLTR-PQNLKDAILK--YNAVYS-----KKWDFTLIDF 577  
Db 46 FELLETVSQWRNCQNLFEYLLKTYKNGYSQFEDLLFKIMGYSEKERFDLPALKSF 105  
QY 578 WDKVLE---EBAQHLYOSILPDWVKIALCLPNICTQIPLLKQKMN-----HSITMSQ 628  
Db 106 YRKMSIIVGEDE-----VLEKLARLVRTKSAAC-----EVLPEKIYRLVGDIESATFSH 154  
QY 629 EQIASLLANAFCTIPPRNNAKMSYSSYPDINFLRPEGRSSRKPEKIKTLFCYFRRVT 688  
Db 155 IQCASLIAMFFSDPTR-----LSFIIILQKTCVAVEKLFELFYFDKMS 200

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: May 26, 2004, 15:28:37 ; Search time 173.315 Seconds  
(without alignments)  
13270.296 Million cell updates/sec

Title: US-09-302-812-1

Perfect score: 7242

Sequence: 1 accggaagtgacgaagcc.....aaatttcattacaaaaa 4070

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 3172214

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ n2p.model -DEV=xlp  
-Q=/cg22\_1/USPTO\_spool\_P/6333148/runat\_26052004\_150052\_5654/app\_query.fasta\_1.12437  
-DB=A Geneseq 29Jan04 -QMT=fastan -SUFFIX=n2p.rag -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blom62 -TRANS=human40.cdi  
-LIST=45 -DOCLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MTN=0 -ALIGN=15  
-MODE=LOCAL -OUTFWT=pt0 -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=6333148 -CGCN 1 1 414 -runat\_26052004\_150052\_5654 -NCPU=6 -ICPU=3  
-NO MAP -LARGEQUERY -NEG SCORES=0 -WAIT -DEPLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : A Geneseq 29Jan04:\*

- 1: Geneseq1980s:\*
- 2: Geneseq1990s:\*
- 3: Geneseq2000s:\*
- 4: Geneseq2001s:\*
- 5: Geneseq2002s:\*
- 6: Geneseq2003as:\*
- 7: Geneseq2003bs:\*
- 8: Geneseq2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	5184	71.6	977	5	Aae25629 Bovine po
2	5184	71.6	977	5	Aau76020 Bovine po
3	5184	71.6	977	5	Aau75799 Bovine po
4	5184	71.6	977	6	Abg72278 Bovine po
5	4635.5	64.0	976	5	Aae25630 Human pol
6	4635.5	64.0	976	5	Aau76021 Human pol
7	4635.5	64.0	976	5	Aau76012 Human pol
8	4635.5	64.0	976	6	Abg72279 Human pol
9	4332.5	59.8	968	5	Aae25631 Murine po
10	4332.5	59.8	968	5	Aau76022 Mouse pol

11	4332.5	59.8	968	5	Aau76013 Mouse pol
12	4332.5	59.8	968	6	Abg72280 Murine p
13	1063.5	14.7	768	4	Abb59491 Drosophil
14	1063.5	14.7	768	4	Aae25632 Fruit fly
15	1063.5	14.7	768	5	Aau76023 Fruit fly
16	1063.5	14.7	768	5	Aau76014 Fruit fly
17	1063.5	14.7	768	6	Abg72281 Fruit fly
18	465	6.4	726	5	Aae25633 Poly aden
19	465	6.4	726	5	Aau76024 Worm poly
20	465	6.4	726	5	Aau76015 Worm poly
21	465	6.4	726	6	Abg72282 C. elegan
22	335.5	4.6	819	4	Abg20721 Novel hum
23	201	2.8	100	4	Abg11103 Novel hum
24	185.5	2.6	954	4	Abb65602 Drosophil
25	184	2.5	2665	4	Aam14533 Peptide #
26	184	2.5	2665	4	Abb33490 Peptide #
27	184	2.5	2665	4	Aam26950 Peptide #
28	184	2.5	2665	4	Abb28314 Human pep
29	184	2.5	2665	4	Abb18950 Protein #
30	184	2.5	2665	4	Aam66665 Human bon
31	184	2.5	2665	4	Aam54270 Human bra
32	184	2.5	2665	4	Abg48336 Human liv
33	184	2.5	2665	4	Aam02259 Peptide #
34	184	2.5	2665	5	Abg36319 Human pep
35	184	2.5	3266	3	Aab42491 Human ORF
36	184	2.5	3371	6	Ab07211 Human p53
37	184	2.5	3664	6	AbR47592 Breast ca
38	184	2.5	3664	6	Ab053027 Human put
39	180	2.5	2194	4	Aam40114 Human pol
40	178	2.5	33	5	Aae25652 Bovine po
41	176	2.4	1210	7	AbR63880 Human AF4
42	175.5	2.4	1163	4	Aau78028 Novel hum
43	174.5	2.4	768	4	Aam78686 Human pro
44	174.5	2.4	962	4	Abg04842 Novel hum
45	174.5	2.4	1462	6	Abp58346 Human cel

# ALIGNMENTS

RESULT 1

Aae25629

ID Aae25629 standard; protein; 977 AA.

XX AC Aae25629;

XX AC Aae25629;

XX 04-NOV-2002 (first entry)

XX Bovine poly adenosine diphosphate-ribose glycohydrolase (bPARG).

Bovine; poly adenosine diphosphate-ribose glycohydrolase; bPARG; enzyme; ADP-ribose; neoplastic disorder; genotoxic oxidative stress; neurotoxic; cardiac disorder; neuronal disorder; reperfusion injury; neurotoxicity; Alzheimer's disease; Huntington's disease; Parkinson's disease; cardiact; cytostatic; vasotropic; neuroprotective; anticonvulsant; gene therapy; antisense therapy.

XX Bos taurus.

XX US6395543-B1.

XX 28-MAY-2002.

XX 23-FEB-2000; 2000US-00511507.

XX 01-MAY-1998; 98US-0083768P.

XX 30-APR-1999; 99US-00302812.

XX (KENT ) UNIV KENTUCKY RES FOUND.

XX Jacobson MK, Jacobson EL, Ame J, Lin W;

XX WPI; 2002-535641/57.

XX N-PSDB; AAD42081.



XX New nucleic acid molecule encoding bovine poly adenosine diphosphate-  
 PT ribose glycohydrolase involved in cellular response to DNA damage,  
 PT inhibition of which is useful for treating neoplastic disorders and  
 PT neurodegenerative diseases.

XX Claim 3; Col 47-45; 77pp; English.

CC The invention relates to an isolated nucleic acid molecule which encodes  
 CC a bovine poly (adenosine diphosphate (ADP)-ribose) glycohydrolase (bPARG)  
 CC which catalyses release of ADP-ribose from an ADP ribose polymer. The  
 CC invention is useful as probes and primer molecules that can be used in  
 CC hybridisation assays and polymerase chain reaction (PCR) amplification.  
 CC The knowledge of the nucleotide sequence of the PARG gene permits the  
 CC preparation of antisense therapeutics containing sequences complementary  
 CC to the mRNA of PARG gene. The antisense therapeutics are useful to treat  
 CC neoplastic disorders and conditions caused by genotoxic oxidative stress  
 CC e.g., cardiac disorders, neuronal disorders, reperfusion injury,  
 CC neurotoxicity, Alzheimer's disease, Huntington's disease and Parkinson's  
 CC disease. The invention is useful in gene therapy and antisense therapy.  
 CC The present sequence is bPARG

XX Sequence 977 AA;

Alignment Scores:

Pred. No.:	0	Length:	977
Score:	5184.00	Matches:	977
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	71.58%	Indels:	0
DB:	5	Gaps:	0

US-09-302-812-1 (1-4070) x AAE25629 (1-977)

QY	258	ATGAGTGGCGGCGCCCGCTGTGAGCCCTGCACCAAGCGACCCCGCTGGAGCGCGCTGCA	317
Db	1	MetSerAlaGlyProGlyCysGluProCysThrLysArgProArgTrpAspAlaAla	20
QY	318	ACTTCTCGCGCGCGCTCGGAGCGCCCGGAGCTTCCCGGCGGAGCGCGGCTCCTC	377
Db	21	ThrSerProAlaAlaSerAspAlaArgSerPheProGlyArgGlnArgValLeu	40
QY	378	GATTCCAGGAGCGTCCGGTGCAGTTCAGGTCCGGCGCTCCTCGTCAGGCTGGCGCTG	437
Db	41	AspSerLysAspAlaProValGlnPheArgValProProSerSerSerGlyCysAlaLeu	60
QY	438	GGCGGCGGCGGACAGCAGCGGAGCGCGCCACTCTCTGTGTTTCAACAGAGACTATA	497
Db	61	GlyArgAlaGlyGlnHisArgGlySerAlaThrSerLeuValPheLysGlnLysThrIle	80
QY	498	ACGATTTGGATGGACACTAAAGGAATCAAGACAGTTCGAATCAGAAAGTTTGCATAGTAAA	557
Db	81	ThrSerTrpMetAspThrLysGlyIleLysThrValGluSerGluSerLeuHisSerLys	100
QY	558	GAATAACAACATCAAGAGAGATCCATCATGATGTTCTGTACAAAAGATACTTTTAT	617
Db	101	GluAsnAsnAsnThrArgGluGluSerMetMetSerSerValGlnLysAspAsnPheTyr	120
QY	618	CAACATAACAATGAAATAATAGAAATGTTTCTCAGCTAGCTTTTGTATAGTCACCAAGTT	677
Db	121	GlnHisAsnMetGluLysLeuGluAsnValSerGlnLeuGlyPheAspLysSerProVal	140
QY	678	GAATAAGGTACAGATTTTGAAGCAGCATCAGATCGCGCTATGTGTAGTGGCAGAAAT	737
Db	141	GluLysGlyThrGlnTyrLeuLysGlnHisGlnThrAlaAlaMetCysLysTrpGlnAsn	160
QY	738	GAAGGCGCACATCAGACGGCTTTGGAAAGTGAACCTCCAGCGGTAACTCTGGTACCA	797
Db	161	GluGlyProHisSerGluArgLeuLeuGluSerMetMetSerValGlnLysAspAsnPheTyr	180
QY	798	GAGCAGTTCAGTAAATGTCATGTCGATCAGTCCGCTCCCAAGAGATCATCACAGTACACA	857
Db	181	GluGlnPheSerAsnAlaAsnValAspGlnSerSerProLysAspHisSerAspThr	200

QY	858	AATAGTGGAGAGTAGAGATAATCAGCAGTGTTCGACATGTAAGCTTCGCAATGCA	917
Db	201	AsnSerGluGluSerArgAspAsnGlnGlnPheLeuThrHisValLysLeuAlaAla	220
QY	918	AAGCAGACGATGGAGATGAACAGCGGCGGAGAGCCAGAGCCAGAGTGTGGCAAG	977
Db	221	LysGlnThrMetGluAspGluGlnGlyArgGluAlaArgSerHisGlnLysCysGlyLys	240
QY	978	GCTTGCCATCTTCGAGAAAGCTGTGACGGGTGTACAGCAGGAGGACAGACAGCTGTGTCC	1037
Db	241	AlaCysHisProAlaGluAlaCysAlaGlyCysGlnGlnGluGluThrAspValValSer	260
QY	1038	GAGAGCCCTTGTCTGGACACCTGGCTCTGAGATGTGCTACTGGACTGAAATCCCAAC	1097
Db	261	GluSerProLeuSerAspThrGlySerGluAspValGlyThrGlyLeuLysAsnAlaAsn	280
QY	1098	AGATTGAATAGACAAAGAGTAGTCTAGGAATCTCTCCATTTGAGAAAGAAAGTGA	1157
Db	281	ArgLeuAsnArgGlnGluSerSerLeuGlyAsnSerProProPheGluLysGluSerGlu	300
QY	1158	CCTCAGTCCACCAATGGATGTAGATAATTCCAAAATAGTTCTCAGGATTCAGAAAGCAGAT	1217
Db	301	ProGluSerProMetAspValAspAsnSerLysAsnSerCysGlnAspSerGluAlaAsp	320
QY	1218	GAAGACAGACAGTCCAGGTTTGTATGAACAGAGATAGCAGTCTCTCAACACAGCAAT	1277
Db	321	GluGluThrSerProGlyPheAspGluGlnGluAspSerSerSerAlaGlnThrAlaAsn	340
QY	1278	AAACCTTCAAGTTCCCAACCAAGAGAGCTGACACTGAGTTGAGGAGCGGCTCTCTGCT	1337
Db	341	LysProSerArgPheGlnProArgGluAlaAspThrGluLeuArgLysArgSerSerAla	360
QY	1338	AAGGAGGTGAGATTGCAATTTCATTTTGAAGGAGGAGAGAGTCCAGCTGGAGATG	1397
Db	361	LysGlyGlyGluIleArgLeuHisPheGlnPheGluGlyGlyGluSerArgAlaGlyMet	380
QY	1398	AATGATGTGAATGCCAAACGACCTGGAGTACTTCTAGCTCTGATGTAGAGTGCAGAAAT	1457
Db	381	AsnAspValAsnAlaLysArgProGlySerThrSerSerLeuAsnValGluCysArgAsn	400
QY	1458	TCTAAGCAACATGGGAGAAAGGATTCATAAATCACAAGATCATTTTCATGAGAGTCCCAAA	1517
Db	401	SerLysGlnHisGlyArgLysAspSerLysIleThrAspHisPheMetArgValProLys	420
QY	1518	GCAGAGCAAAAGAAAGAACAAATGTGAATCAAAATCAAAACATCAAAAGAGAGATC	1577
Db	421	AlaGluAspLysArgLysGluGlnCysGluMetLysHisGlnArgThrGluArgLysIle	440
QY	1578	CCTAAATACATTCACCTCCTCCCTTCTCCAGATAAGAAATGGCTTGGACTCTTATGAG	1637
Db	441	ProLysTyrIleProProHisLysLeuSerProAspLysLysTrpLeuGlyThrProLysGlu	460
QY	1638	GAGATCAGGAGAAATGCAAGGTGTGGGATCCGCTCCCTCCCTTTCAGACCATCTGCAAT	1697
Db	461	GluMetArgArgMetProArgCysGlyIleArgLeuProProLeuArgProSerAlaAsn	480
QY	1698	CACAGAGTCACTATTCGGGTAGATCTTTTGGAGATAGAGAGTTCCTAAACCTTTCCCA	1757
Db	481	HisThrValThrIleArgValAspLeuLeuArgGlyGluValProLysProPhePro	500
QY	1758	ACACATTTTAAAGATTGTGGCACAACAGCATGTTCAGATGCTTGTTCAGACAAAC	1817
Db	501	ThrHisPheLysAspLeuTrpAspAsnLysHisValLysMetProCysSerGluGlnAsn	520
QY	1818	TTGTACCTCTGCGAAGATGAGATGGTGCAGCGAGCTGCAGGCGCGGCTGGGAACTCAT	1877
Db	521	LeuTyrProValGluAspGluAsnGlyGluArgAlaAlaGlySerArgTrpGluLeuIle	540
QY	1878	CAGACTGCATCTCTCAACAGGCTACTCGGCCCCAGACCTGAGAGATGCTTATTCGAG	1937
Db	541	GlnThrAlaLeuLeuAsnArgLeuThrArgProGlnAsnLeuLysAspAlaIleLeuLys	560

1938	TACAAATGGGCATATTTCTAAGAAATGGCACTTTACAGCTTTGATTGATTTCTGGATAAG	1997	QY
561	TyrAenValAlaTyrSerLysLysTrpAspPheThrAlaLeuIleAspPheTrpAspLys	580	DB
1998	GTACTAGAAGACAGAGAGCTCAACACTTGTATCAGTCCATCTTGGCTGTATATCGTGA	2057	QY
581	ValLeuGluGluAlaGluAlaGlnHisLeuTyrGlnSerIleLeuProaspMetValLys	600	DB
2058	ATTGCACCTCTGTCTGCCAAATATTTGTATCCAGGCAATACCACTCTCTGMAACAGAGATG	2117	QY
601	IleAlaLeuCysLeuProAsnIleCysThrGlnProIleProLeuLeuLysGlnLysMet	620	DB
2118	AATCAATTCATCACAATGTCACAGAAACAGATTGCCAGTCTTTAGCTTAATGCTTCTTC	2177	QY
621	AsnHisSerIleThrMetSerGlnGluGlnIleAlaSerLeuLeuAlaAsnAlaPhePhe	640	DB
2178	TGCACGTTTCCACGACGCAATGCCAAGATGAATCAGAGATATCCAGTATTCCAGATATT	2237	QY
641	CysThrPheProArgArgAsnAlaLysMetLysSerGluTyrSerTyrProAspIle	660	DB
2238	AACTTCAATCGGTTGTTGAAGAGCGTTTCATCAAGAAACACAGAGAGCTTAAACCGTC	2297	QY
661	AsnPheAsnArgLeuPheGluGlyArgSerSerArgLysProGluLeuLysThrLeu	680	DB
2298	TTCTGCTACTTTAGAGACTCACAGAGAAACCACTGGTTCGTGACATTTCAAGA	2357	QY
681	PheCysTyrPheArgArgValThrGluLysLysProThrGlyLeuValThrPheThrArg	700	DB
2358	CAGAGCTTTGAAGATTTCCAGAGTGGGAAAGATGTGAAAACTCTCTGACTCGACTGAT	2417	QY
701	GlnSerLeuGluAspPheProGluTrpGluArgCysGluLysLeuLeuThrArgLeuHis	720	DB
2418	GTCACTTAGAAGGTACATAGAGAAACGGCCAGGCATCTCACAGTGCATTTTGCA	2477	QY
721	ValThrTyrGluGlyThrIleGluGlyAsnGlyGlnGlyMetLeuGlnValAspPheAla	740	DB
2478	AAACCGTTTCGTTGGAGGTGGTGAACACAGTGCAGGACTTGTGCAAGAGAAATCCGTTT	2537	QY
741	AsnArgPheValGlyGlyValThrSerAlaGlyLeuValGlnGluGluIleArgPhe	760	DB
2538	TTAATCAACCTCAGTTGATTGTTTTCACGCTCTTTCACGTGAGTGTGGATCACATGAA	2597	QY
761	LeuIleAsnProGluLeuIleValSerArgLeuPheThrGluValLeuAspHisAsnGlu	780	DB
2598	TGTCCTTATCATCACAGTACTGAGCAGTACAGTGAATACACAGCTATCCGAAACATAC	2657	QY
781	CysLeuIleIleThrGlyThrGluGlnTyrSerGluTyrThrGlyTyrAlaGluThrTyr	800	DB
2658	CGCTGGCCCGGACCATGAAGACAGAGCGAAAGGAGCAGCTGGCAGAGCGCCACGACT	2717	QY
801	ArgTrpAlaArgSerHisGluAspArgSerGluArgAspAspTrpGlnArgArgThrThr	820	DB
2718	GAGATCGTCGCATCGACGCCCTCCACTTCAGACGCTACCTCCACAGTTGTGGCCGAG	2777	QY
821	GluIleValAlaIleAspAlaLeuHisPheArgArgTyrLeuAspGlnPheValProGlu	840	DB
2778	AAGATCAGACGGAGCTTAAACAGGCCTTACTGTGGAATTTCTCGTCTCGAGTTTCTTCA	2837	QY
841	LysIleArgArgGluLeuAsnLysAlaTyrCysGlyPheLeuArgProGlyValSerSer	860	DB
2838	GAGAACCTGTCTCAGTGGCTACAGGAAACTGGGGCTGTGGTGCCTTTGGGGTGTGCT	2897	QY
861	GluAsnLeuSerAlaValAlaThrGlyAsnTrpGlyCysGlyAlaPheGlyGlyAspAla	880	DB
2898	AGACTAAAGCCTTAATACAGATCCTCGCAGCTCTGTAGCTGACGACGCTGTGTTAT	2957	QY
881	ArgLeuLysAlaLeuIleGlnIleLeuAlaAlaValAlaGluArgAspValValTyr	900	DB
2958	TTCACTTTGGGGACTCAGAACTCATGAGACACATTTTACAGCATGCAATTCCTCACT	3017	QY
901	PheThrPheGlyAspSerGluLeuMetArgAspIleTyrSerMetHisThrPheLeuThr	920	DB
3018	GAGAGGAAACTGACTGTTGGAGAGATATATAAGCTGTGCTACGATATTACAATGAAGA	3077	QY

Db	921	GIuAGLySLeuThrValGlyGluValTyrLysLeuLeuLeuArgTyrTyrAsnGluGlu	940
Qy	3078	TCGAGAAATCGTCACCCCGGACGACATCAAGCTTTATCATTTATATACCATGCA	3137
Db	941	CysArgAsnCysSerThrProGlyProAspIleGlyLeuTyrProPheIleTyrHisAla	960
Qy	3138	GTTGAGTCCTGTACACAGACACCAACCGCGGACAAAGACGGGGCC	3188
Db	961	ValGluSerCysThrGlnThrThrAsnGlnProGlyGlnArgThrGlyAla	977
RESULT 2			
AAU76020			
ID	AAU76020	standard; protein; 977 AA.	
XX	AAU76020;		
XX	08-MAY-2002	(first entry)	
XX		Bovine poly (ADP-ribose) glycohydrolase (PARG) protein sequence.	
XX		Cow; poly (ADP-ribose) glycohydrolase; PARG; DNA repair; ADP-ribose;	
KW		adenosine diphosphate-ribose; neoplastic disorder; adenocarcinoma;	
KW		leukaemia; lymphoma; melanoma; myeloma; sarcoma; teratocarcinoma;	
KW		hyperplasia; hypertrophy; reperfusion; ischaemia; heart attack; stroke;	
KW		neurodegenerative disease; neurological disorder; Alzheimer's disease;	
KW		Huntington's disease; Parkinson's disease.	
XX			
OS		Bos taurus.	
XX			
FH	Key	Location/Qualifiers	
FT	Region	601. .617	
FT		/note= "Represents PARG oligopeptide #1"	
FT	Region	761. .770	
FT		/note= "Represents PARG oligopeptide #2"	
FT	Region	771. .801	
FT		/note= "Represents PARG oligopeptide #3"	
FT	Region	849. .880	
FT		/note= "Represents PARG oligopeptide #4"	
XX			
PN	US6337202-B1.		
XX			
PD	08-JAN-2002.		
XX			
PF	23-FEB-2000; 2000US-00511477.		
XX			
PR	01-MAY-1998; 98US-0083768P.		
PR	30-APR-1999; 99US-00302812.		
XX			
PA	(KENT ) UNIV KENTUCKY RES FOUND.		
XX			
PI	Jacobson MK, Jacobson EL, Ame J, Lin W;		
XX			
DR	WPI; 2002-163240/21.		
DR	N-PSDB; ABK14931.		
XX			
PT	Novel isolated and purified poly (ADP-ribose) glycohydrolase protein		
PT	which catalyses release of ADP-ribose from ADP ribose polymer, useful for		
PT	treating neoplastic and neurological disorders, heart attack and stroke.		
XX			
PS	Claim 2; Col 47-52; 81pp; English.		
XX			
CC	The present invention relates to a new poly (ADP-ribose) glycohydrolase		
CC	(PARG) protein which catalyses release of ADP-ribose from an ADP		
CC	(adenosine diphosphate)-ribose polymer. The PARG molecule of the		
CC	invention is useful for generating antibodies and can be inhibited or		
CC	activated for diagnosing and treating neoplastic disorders such as		
CC	adenocarcinoma, leukaemia, lymphoma, melanoma, myeloma, sarcoma,		
CC	teratocarcinoma, hyperplasia and hypertrophy, reperfusion following		
CC	ischaemia, heart attack, stroke, neurodegenerative diseases, neurological		
CC	disorders including Alzheimer's, Huntington's and Parkinson's diseases,		
CC	and related conditions. PARG levels may be enhanced to suppress DNA		
CC	repair and increase the cell's susceptibility to chemotherapy drugs.		

CC Antagonists of PARC are administered to treat or prevent neoplastic  
 CC disorders. The present amino acid sequence represents the bovine PARC  
 CC protein of the invention. This protein is one of several PARC proteins  
 CC (AAU76020-AAU76024) of the invention

XX  
 SQ Sequence 977 AA;

Alignment Scores:  
 Pred. No.: 0 Length: 977  
 Score: 5184.00 Matches: 977  
 Percent Similarity: 100.00% Conservative: 0  
 Best local Similarity: 100.00% Mismatches: 0  
 Query Match: 71.58% Indels: 0  
 DB: 5 Gaps: 0

US-09-302-812-1 (1-4070) x AAU76020 (1-977)

QY 258 ATGAGTGGCGGCGCGCTGTGAGCCCTGCACCAAGCGACCCCGCTGGGAGCGCGCTGCA 317  
 DB 1 MetSerAlaGlyProGlyCysGluProCysThrLysArgProArgTrpAspAlaAla 20  
 QY 318 ACTTCTCCCGCGCGCTCGGACGCGCGAGCTTCCCGGAGGAGGCGCGCTCTC 377  
 DB 21 ThrSerProAlaAlaSerAspAlaArgSerPheProGlyArgGlnArgValLeu 40  
 QY 378 GATTCAAGACGCTCCGCTGAGTTCAGGTCCTCCGCGCTCTCTCTGTCAGGCTGCGCCCTG 437  
 DB 41 AspSerLysAspAlaProValGlnPheArgValProProSerSerSerGlyCysAlaLeu 60  
 QY 438 GCGCGCGCGGACAGCACCGGCGCGCCACCTCTCTGTTTCAACAGAGACTATA 497  
 DB 61 GlyArgAlaGlyGlnHisArgGlySerAlaThrSerLeuValPheLysGlnLysThrIle 80  
 QY 498 ACCAGTGTGATGACACTAAGGAATCAAGACAGATGTAATCAAGAAAGTTTGATAGTAA 557  
 DB 81 ThrSerTrpMetAspThrLysGlyIleLysThrValGluSerGluSerLeuHisSerLys 100  
 QY 558 GAAACACAAATACAGAGAGAAATCCATGATGAGTCTCTGACAAAGAAATCACTTTAT 617  
 DB 101 GluAsnAsnThrArgGluGluSerMetMetSerValGlnLysAspAsnPheTyr 120  
 QY 618 CAACATAACATGAAAAATAGAAAATGTTCTCAGCTAGTGGTTTGTAAAGTCAACAGTT 677  
 DB 121 GlnHisAsnMetGlnLysLeuGluAsnValSerGlnLeuGlyPheAspLysSerProVal 140  
 QY 678 GAAAAAGGTACACAGTATTGAAGACGATCAGACTCGGCTATGTGAAGTGGAGAT 737  
 DB 141 GluLysGlyThrGlnTyrLeuLysGlnHisGlnThrAlaAlaMetCysLysTrpGlnAsn 160  
 QY 738 GAAGGCCACACTCAGAACGGCTTTGGAAAGTGAACCTCCAGCGTAACTCTGGTACCA 797  
 DB 161 GluGlyProHisSerGluArgLeuLeuGluSerGluProAlaValThrLeuValPro 180  
 QY 798 GAGCAGTTTCAATGTAATGCTAGTCACTCGCTCCCAAGGATGATCAGTGCACCA 857  
 DB 181 GluGlnPheSerAsnAlaAsnValAspGlnSerSerProLysAspAspHisSerAspThr 200  
 QY 858 AATAGTGAGAGAGTAGAGATAATCAGCAGTGTGACATGTAAGCTTGGGAATGCA 917  
 DB 201 AsnSerGluGluSerArgAspAsnGlnGlnPheLeuThrHisValLysLeuAlaAsnAla 220  
 QY 918 AGCAGACGATGGAGATGAACGGGCGAGAGAGCCAGAGAGCCACCAAGAGTGTGGCAAG 977  
 DB 221 LysGlnThrMetGluAspGluGlnGlyArgGluAlaArgSerHisGlnLysCysGlyLys 240  
 QY 978 GCTTGCCATCTCGCAGAGCCTGTGCGAGGCTGTGAGGAGGAGGAGACAGACGCTGGTCC 1037  
 DB 241 AlaCysHisProAlaGluAlaCysAlaGlyCysGlnGlnGluGluThrAspValValSer 260  
 QY 1038 GAGAGCCCTTGTGCGACACTGGCTCTGAGATGTTGGTACTGACGCTGAAATGCCAAC 1097  
 DB 261 GluSerProLeuSerAspThrGlySerGluAspValGlyThrGlyLeuLysAsnAlaAsn 280

QY 1098 AGATTGAATAGACAAAGAAAGTAGTCTAGGAAATTTCTCTCCATTTGAGAAAGAAAGTGAA 1157  
 DB 281 ArgLeuAsnArgGlnGluSerSerLeuGlyAsnSerProProPheGluLysGluSerGlu 300  
 QY 1158 CCTGAGTCACCAATGGATGTAGATAATTCAAAAATAGTTGTCTCAGGATTCAGAACACAT 1217  
 DB 301 ProGluSerProMetAspValAspAsnSerLysAsnSerCysGlnAspSerGluAlaAsp 320  
 QY 1218 GAAGACCAAGTCCAGGTTTTGATGAACAGAGAGATAGCAGTTCTGTCTCAACACAGCAAT 1277  
 DB 321 GluGluThrSerProGlyPheAspGluGlnGluAspSerSerSerAlaGlnThrAlaAsn 340  
 QY 1278 AAACCTTCACAGGTTCCCAACCAAGAGAGCTGACACTGAGTTTCAGGAACGGTCTCTGCT 1337  
 DB 341 LysProSerArgPheGlnProArgGluAlaAspThrGluLeuArgLysArgSerSerAla 360  
 QY 1338 AAGGAGGTGAGATTCGATTACATTTCCAAATTTAGAGGAGGAGAGTTCGAGCTGGAATG 1397  
 DB 361 LysGlyGlyGluIleArgLeuHiePheGlnPheGluGlyGlyLysArgAlaGlyMet 380  
 QY 1398 AATGATGTGAATGCCAAACGACCTGGAAGTACTTCTAGCCTGAATGTAGAGTGCAGAAAT 1457  
 DB 381 AsnAspValAsnAlaLysArgProGlySerThrSerSerLeuAsnValGluCysArgAsn 400  
 QY 1458 TCTAAGCAACATGGGAGAAAGGATTTCTAAATTCACAGATCATTTTCATGAGAGTCCCAA 1517  
 DB 401 SerLysGlnHisGlyArgLysAspSerLysIleThrAspHisPheMetArgValProLys 420  
 QY 1518 GCAGAGGCAAAAGAAAAGAACATGTGAATGAACATCAAGACACAGAGAGGAGATC 1577  
 DB 421 AlaGluAspLysArgLysGluGlnCysGluMetLysHisGlnArgThrGluArgLysIle 440  
 QY 1578 CCTAATATCATTTCCACCTCACCTTTCTCAGATAAGAAATGGCTTGGAACTCTATTGAG 1637  
 DB 441 ProLysTyrIleProProHisLeuSerProAspLysLysTrpLeuGlyThrProIleGlu 460  
 QY 1638 GAGATGAGGAGAAATCCCAAGGTGTGGATCCGGCTGCTCCCTTGAGACCATCTGCCAAT 1697  
 DB 461 GluMetArgArgMetProArgCysGlyIleArgLeuProProLeuArgProSerAlaAsn 480  
 QY 1698 CACACAGTGACTATTTCGGGTAGATCTTTTGGCAATAGAGAAAGTTCTCTAAACCTTTCCCA 1757  
 DB 481 HisThrValThrIleArgValAspLeuLeuArgIleGlyGluValProLysProPhePro 500  
 QY 1758 ACACATTTTAAAGATTTGTGGCAACCAACAGCATGTTTAAGATCCTTGTTCAGAACAAAC 1817  
 DB 501 ThrHisPheLysAspLeuTrpAspAsnLysHisValLysMetProCysSerGluGlnAsn 520  
 QY 1818 TTGTACCTGTGGAAGATGAGAAATGGGACTTTACAGCTTTTGTATGCTTTCTGGGATAG 1877  
 DB 521 LeuTyrProValGluAspGluAsnGlyGluArgAlaAlaGlySerArgTrpGluLeuIle 540  
 QY 1878 CAGACTGCTCTTCCAAACAGGCTCACTCGGCCCCAGAACCTCAAGGATCTATTCTTGAAG 1937  
 DB 541 GlnThrAlaLeuLeuAsnArgLeuThrArgProGlnAsnLeuLysAspAlaIleLeuLys 560  
 QY 1938 TACAATGTGGCATATTCTAAGAAATGGGACTTTACAGCTTTTGTATGCTTTCTGGGATAG 1997  
 DB 561 TyrAsnValAlaTyrSerLysLysTrpAspPheThrAlaLeuIleAspPheTrpAspLys 580  
 QY 1998 GTACTAGAAAGACAGACAGCTCAACACTTTGTATCAGTCCATCTTGCCTCATATGGTAAA 2057  
 DB 581 ValLeuGluGluAlaGluAlaGlnHisLeuTyrGlnSerIleLeuProAspMetValLys 600  
 QY 2058 ATTGCACTCTGTCTGCCAAATTTTGTACCCAGCCCAATACCACCTCTGTAACACAGAGATG 2117  
 DB 601 IleAlaLeuCysLeuProAsnIleCysThrGlnProIleProLeuLeuLysGlnLysMet 620  
 QY 2118 AATCATTTCATCAATGTACAGGAAACAGATGTCAGCTCTTTTACGATGCTTTCTTC 2177  
 DB 621 AsnHisSerIleThrMetSerGlnGluGlnIleAlaSerLeuLeuAlaAsnAlaPhePhe 640  
 QY 2178 TGCACTGTTCCACGACGAAATGCCAAGATGAATCAGAGTATTTCAGTATTCCAGATATT 2237

641 CysThrPheProArgAsnAlaLysMetLysSerGluTyrSerSerTyrProaspIle 660  
 2238 AACTTCATCGTCTGTTTGAAGAGCGTTTCATCAAGGAAACACAGAAAGCTTAAACGCTC 2297  
 661 AsnPheAsnA-gLeuPheGluGlyArgSerSerA-gLysProGluLysLeuLysThrLeu 680  
 2298 TTCTGCTACTTTAGAGAGTACAGAGAAACCAACCACTGGTGGTGGTACATTCACAGA 2357  
 681 PheCysTyrPheArgArgValThrGluLysLysProThrGlyLeuValThrPheThrArg 700  
 2358 CAGAGCTTTGAAGATTTCCAGAGTGGGAAAGATGTGAATACTCTGACTCGACTGCAT 2417  
 701 GlnSerLeuGluAspPheProGluTrpGluA-gCysGluLysLeuLeuThrArgLeuHis 720  
 2418 GTCACCTTACGAAGTACCATAGAGAAACCGCCAGCGCATCTACAGTGGATTTGCA 2477  
 721 ValThrTyrGluGlyThrLeuGluLysAsnGlyGlnGlyMetLeuGlnValAspPheAla 740  
 2478 AACCGTTTCTGGAGTGTGTAAACAGTGCAGGACTTGTGCAGAGAAATCCGCTTT 2537  
 741 AsnArgPheValGlyGlyValThrSerAlaGlyLeuValGlnGluLeuArgPhe 760  
 2538 TTAATCAACCTCAGTGTATTGTTTCAAGCTCTTCACTGAGTGTGTGATCAATGAA 2597  
 761 LeuLeuAsnProGluLeuValSerA-gLeuPheThrGluValLeuAspHisAsnGlu 780  
 2598 TGTCTTATCATCAGCTACTGAGCAGTACAGTGAATACAGGCTATCGGAAACATAC 2657  
 781 CysLeuLeuLeuThrGlyThrGluGlnTyrSerGluTyrThrGlyTyrAlaGluThrTyr 800  
 2658 CGCTGGCGCCGAGCCATGAAGACAGAGCGAAGGAGCACTGGCAGAGGCCACGACT 2717  
 801 ArgTyrAlaArgSerHisGluAspArgSerGluArgAspTrpGlnArgArgThrThr 820  
 2718 GAGATCGTGCCTATGAGCGCCCTCCACTTCAGAGCTTACCTCGACAGTTTGGCCGAG 2777  
 821 GluLeuValAlaLeuAspAlaLeuHisPheArgArgTyrLeuAspGlnPheValProGlu 840  
 2778 AAGATCAGCGGAGCTTAACAGGCTTACTGTGGATTCTTCTGCTGGAGTTTCTTCA 2837  
 841 LysIleArgArgGluLeuAsnLysAlaTyrCysGlyPheLeuArgProGlyValSerSer 860  
 2838 GAGAACCTGTCTGAGTGGCTACAGAAACTGGGCGTGTGCTGCTTGGGGTGATGCT 2897  
 861 GluAsnLeuSerAlaValAlaThrGlyAsnTrpGlyCysGlyAlaPheGlyGlyAspAla 880  
 2898 AGACTAAAGCCTTAATACAGATCTCGCAGCTGCTGTAGTGCAGCAGACGTGGTTTAT 2957  
 881 ArgLeuLysAlaLeuLeuGlnIleAlaAlaAlaValAlaGluArgAspValValTyr 900  
 2958 TTCACCTTTGGGGACTCAGAACTGATGAGAGACATTTACAGCATGCATACATTCCTCACT 3017  
 901 PheThrPheGlyAspSerGluLeuMetArgAspIleTyrSerMetHisThrPheLeuThr 920  
 3018 GAGAGGAACCTGACTGTGGAGAGATATATAGCTGTGCTGCTACCATATTACAAATGAAGA 3077  
 921 GluArgLysLeuThrValGlyGluValTyrLysLeuLeuLeuArgTyrTyrAsnGluGlu 940  
 3078 TGCAGAACTGCTCCACCCCGGACACATCAAGCTTTATCCATTCATATACATGCA 3137  
 941 CysArgAsnCysSerThrProGlyProAspIleLysLeuTyrProPheIleTyrHisAla 960  
 3138 GTTAGTGTCTGTACACAGACACCAACCGCGGACCAAGAGCGGGGCC 3188  
 961 ValGluSerCysThrGlnThrThrAsnGlnProGlyGlnArgThrGlyAla 977

08-MAY-2002 (first entry)  
 Bovine poly (ADP-ribose) glycohydrolase (PARG) protein sequence.  
 Cow; poly (ADP-ribose) glycohydrolase; PARG; PARG modulator; ADP-ribose;  
 adenosine diphosphate-ribose; DNA repair; apoptosis; neoplasia; cancer;  
 inherited genetic disease; myocardial infarction; vascular stroke; aging;  
 neurodegeneration; Huntington's disease; Parkinson's disease;  
 Alzheimer's disease; neurotoxicity.  
 Bos taurus.  
 Key Location/Qualifiers  
 Region 601..617 /note= "Represents PARG oligopeptide #1"  
 Region 761..770 /note= "Represents PARG oligopeptide #2"  
 Region 771..801 /note= "Represents PARG oligopeptide #3"  
 Region 849..880 /note= "Represents PARG oligopeptide #4"  
 US6333148-B1.  
 25-DEC-2001. 99US-00302812.  
 30-APR-1999; 99US-0083768P.  
 01-MAY-1998; 98US-0083768P.  
 (KENT) UNIV KENTUCKY RES FOUND.  
 Jacobson MK, Jacobson EL, Ame J, Lin W;  
 WPI; 2002-153820/20.  
 N-PSDB; ABK14493.  
 Screening compounds for modulation of poly (ADP-ribose) glycohydrolase,  
 useful potentially for treating diseases associated with DNA damage, e.g.  
 cancer.  
 Claim 3; Col 45-52; 80pp; English.  
 The present invention relates to a new method for screening compounds for  
 ability to modulate activity of an enzyme that hydrolyses ADP (adenosine  
 diphosphate)-ribose from an ADP-ribose polymer. The compounds are  
 inhibitors or activators of PARG (poly (ADP-ribose) glycohydrolase) and  
 are used to treat or prevent any condition associated with DNA damage,  
 e.g. neoplasia, inherited genetic diseases, myocardial infarction,  
 vascular stroke, aging and neurodegeneration e.g. Huntington's,  
 Parkinson's or Alzheimer's diseases, or neurotoxicity generally.  
 Compounds identified by the new method are more effective than known  
 inhibitors and have fewer side effects. The present amino acid sequence  
 represents the bovine PARG protein of the invention. This protein is one  
 of several PARG proteins (AAU75799 and AAU76012-AAU76015) of the  
 invention  
 Sequence 977 AA;  
 Alignment Scores:  
 Pred. NO.: 0 Length: 977  
 Score: 5184.00 Matches: 977  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 71.58% Indels: 0  
 DB: 5 Gaps: 0  
 US-09-302-812-1 (1-4070) x AAU75799 (1-977)  
 Qy 258 ATGAGTGGGCGCCCGCTGTGAGCTTCACCAAGCGACCCCGCTGGACCGCTGCA 317  
 Db 1 MetSerAlaGlyProGlyCysGluProCysThrLysArgProArgTyrAspAlaAla 20

318 ACTTCTCCGCGCGCTCCGACGCGCGGAGCTTCCCGCGCAGCGAGCGGCTCCTC 377  
21 ThrSerProAlaAlaSerAspAlaArgSerPheProGlyAArgGlnArgValIleu 40  
378 GATTCCAGGACGCTCCCGTGCAGTTTCAAGGTCCCGCGCTCCTCGTCAGGCTGGCGCCCTG 437  
41 AspSerLysAspAlaProValGlnPheArgValProProSerSerSerGlyCysAlaLeu 60  
438 GCGCGCGCGGACGACGACCGAGGCGCGCCCTCTCTTGTGTTTCAACAGAGACTATA 497  
61 GlyArgAlaGlyGlnHisArgGlySerAlaThrSerLeuValPheLysGlnLysThrIle 80  
498 ACCAGTTGAGTGGACACTAAAGGAATCAAGCAGTTGAATCAGAAAGTTTCATAGTAAA 557  
81 ThrSerTrpMetAspThrLysGlyLysThrValGluSerGluSerLysLeuHisSerLys 100  
558 GAAACCAACAATACAGAGAGAATCCATGATGAGTTCTGTACAAAAGATAACTTTTAT 617  
101 GluAsnAsnThrArgGluSerMetMetSerValGlnLysAspAsnPheTyr 120  
618 CAACATAACATGGAAAAATTAGAAATGTTTCTCAGCTAGGTTTGTATAGTCACAGTT 677  
121 GlnHisAsnMetGluLysLeuGluAsnValSerGlnLeuGlyPheAspLysSerProVal 140  
678 GAAAGAGTTACACAGTATTTGAAGCAGCATCAGACTGCGGCTATCTGTAAAGTGGCAGAA 737  
141 GluLysGlyThrGlnTyrLysLysGlnHisGlnThrAlaAlaMetCysLysTrpGlnAsn 160  
738 GAAGGCGCACACTCAGAACCGGCTTTTGGAAAGTGAACCTCCAGCGGTAACCTCTGTACCA 797  
161 GluGlyProHisSerGluArgLeuLeuGluSerGluProAlaValThrLeuValPro 180  
798 GAGCAGTTCAATGCTAATGTCATCATCTGCTGCTCCCAAGGATGATCAGTGACACA 857  
181 GluGlnPheSerAsnAlaAsnValAspGlnSerSerProLysAspAspHisSerAspThr 200  
858 AATAGTGGAGGAGTAGAGATAATCAGCAGTTTTTGGACATGTAAGCTTCGCAATGCA 917  
201 AsnSerGluLysArgAspAsnGlnPheLeuThrHisValLysLeuAlaAsnAla 220  
918 AAGCAGCAGTGAAGATGAACAGGCGCAGAGAGCCAGAGCCACAGAGTGTGGCAAG 977  
221 LysGlnThrMetGluAspGluGlnArgGluAlaArgSerHisGlnLysCysGlyLys 240  
978 GCTTCCCATCTCGCAGGCTGTCAGGCTGTCAGCAGGAGCAGACGTGTGTCC 1037  
241 AlaCysHisProAlaGluAlaCysAlaGlyCysGlnGlnGluLysThrAspValValSer 260  
1038 GAGAGCCCTCTCGGACACTGGCTCTGAGGATGTTGTGTAAGTCACTGAAATGCCAAC 1097  
261 GluSerProLeuSerAspThrGlySerGluAspValGlyThrGlyLeuLysAsnAlaAsn 280  
1098 AGATTGAATAGACAGAAAGTAGTCTAGAAATTTCTCCATTTTCAGAAAGAAAGTGA 1157  
281 ArgLeuAsnArgGlnGluSerSerLeuGlyAsnSerProPheGluLysGluSerGlu 300  
1158 CTGTAGTCCCAATGATGATGATATATCCAAAAATAGTTGTGAGATTTCAGAGCAGAT 1217  
301 ProGluSerProMetAspValAspAsnSerLysAsnSerCysGlnAspSerGluAlaAsp 320  
1218 GAAGAGACAAGTCCAGGTTTGTATGAACAGGAGATAGCAGTTCTGCTCAAAACAGCAAT 1277  
321 GluGluThrSerProGlyPheAspGluGlnGluAspSerSerAlaGlnThrAlaAsn 340  
1278 AAACCTTCAGGTTTCAACCAAGAGAGCTGACACTGAGTTGAGGAGCGGCTCTGCT 1337  
341 LysProSerArgPheGlnProArgGluAlaAspThrGluLeuArgLysArgSerAla 360  
1338 AAGGAGGTGAGATTCGATTACATTTCCAAATTTGAAGGAGGAGAGAGTCCAGCTGGAATG 1397  
361 LysGlyGlyGluLysLeuHisPheGlnPheGluGlyGlyLysSerArgAlaGlyMet 380  
1398 AATGATGTGAATGCCAAACGACCTGGAAGTACTTCTAGCCGTAATGTAGATGTCAGAAAT 1457

381 AsnAspValAsnAlaLysArgProGlySerThrSerSerLeuAsnValGluCysArgAsn 400  
1458 TCTAAGCAACATGGAGAAAGGATTTAAATTCACAGATCATTTTCATGAGAGTGCCCAA 1517  
401 SerLysGlnHisGlyArgLysAspSerLysIleThrAspHisPheMetArgValProLys 420  
1518 GCAGAGCAACAAAGAAAGACAAATGTGAATGAACATCAAGAAACAGAAAGGAAGATC 1577  
421 AlaGluAspLysArgLysGluGlnCysGluMetLysHisGlnArgThrGluArgLysIle 440  
1578 CCTAAATACATTTCCACCTCCTTCTCCAGATGAAGAAATGGCTTGGAACTCCTATTGAG 1637  
441 ProLysTyrIleProProHisLeuSerProAspLysLysTrpLeuGlyThrProIleLeu 460  
1638 GAGATGAGGAGAAATGCCAGGTGGGATCCGGCTGCTCCTTGGAGCCATCTGCCAAT 1697  
461 GluMetArgArgMetProArgCysGlyIleArgLeuProProLeuArgProSerAlaAsn 480  
1698 CACACAGTACTATTCCGGTAGATCTTTTGGGAATAGAGAAATTCCTAAACCTTTCCCA 1757  
481 HisThrValThrIleArgValAspLeuLeuArgIleGlyGluValProLysProPhePro 500  
1758 ACACATTTTAAAGATTTTGGGACAAACAGCAATGTTAAGATGCTTGTTCAGACAAAC 1817  
501 ThrHisPheLysAspLeuTrpAspAsnLysHisValLysMetProCysSerGluGlnAsn 520  
1818 TTGTACCTGTGGAGATGAGATGGTGGAGAGCTCAGCAGCCGCGTGGGAACTCAT 1877  
521 LeuTyrProValGluAspGluAsnGlyGluAlaGlySerArgTrpGluLeuIle 540  
1878 CAGACTGCACCTTCAACAGGCTCACTCGGCGCCAGAACCTGAAAGGATGCTATTCTGAAG 1937  
541 GlnThrAlaLeuLeuAsnArgLeuThrArgProGlnAsnLeuLysAspAlaIleLeuLys 560  
1938 TACAATGTGGCATATCTTAAGAAATGGGACTTTACAGCTTTTGTATGATTTCTGGGATAG 1997  
561 TyrAsnValAlaTyrSerLysLysTrpAspPheThrAlaLeuIleAspPheTrpAspLys 580  
1998 GTACTAGAGAGAGCAGAGCTCAACACTGTATCAGTCCATCTTGTCCCTGATATGTTGAAA 2057  
581 ValLeuGluGluAlaGluAlaGlnHisLeuLysGlnSerIleLeuProAspMetValLys 600  
2058 ATTGCACCTCTCTCGCAATATTTGTACCCAGCAATATCACCTCCTGAAACAGAGATG 2117  
601 IleAlaLeuCysLeuProAsnIleCysThrGlnProIleProLeuLeuLysGlnLysMet 620  
2118 AATCATTTCCATCACAATGTCACAGGAACAGATTCCTAGTCTTTTAGCTAATGCTTCTTC 2177  
621 AsnHisSerIleThrMetSerGlnGluGlnIleAlaSerLeuLeuAlaAsnAlaPhePhe 640  
2178 TGCAGCTTTCCAGCAGCAATGCCAAGATGAATCAGATGATTCAGTATTCAGATATT 2237  
641 CysThrPheProArgArgAsnAlaLysMetLysSerGluTyrSerSerTyrProAspIle 660  
2238 AACTTCAATCGGTTGTTTGAAGGAGCTTCATCAAGAAACAGAGAACTTAAACGCTC 2297  
661 AsnPheAsnArgLeuPheGluGlyArgSerSerArgLysProGluLysLeuLysThrLeu 680  
2298 TTCTCTACTTTTAGAGAGTACAGAGAAACCCACTGGTGTGTCACATTCACAAGA 2357  
681 PheCysTyrPheArgArgValThrGluLysLysProThrGlyLeuValThrPheThrArg 700  
2358 CAGAGCTTTGAAGATTTTCCAGAGTGGGAAAGATGTGAAAACTCCTCACTCGACTGAT 2417  
701 GlnSerLeuLeuAspPheProGluTrpGluArgCysGluLysLeuLeuThrArgLeuHis 720  
2418 GTCACTTACAGAGGTACCATAGAGAAACCGCCAGGCGCATGCTACAGGTGATTTTCCA 2477  
721 ValThrTyrGluGlyThrIleGluGlyAsnGlyGlnGlyMetLeuGlnValAspPheAla 740  
2478 AACCCCTTTCGTTGGAGTGGTGTACACAGTGCAGGACTTGTGCAAGAAAGAAATCCGCTTT 2537

Db 741 AsnArgPheValGlyGlyValThrSerAlaGlyLeuValGlnGluLeuArgPhe 760  
 QY 2538 TTAATCAACCCCTGAGTTGATTTTCAAGCTCTTCACTGAGGTGCTGGATCAACAATGAA 2597  
 Db 761 LeuIleAsnProGluLeuValSerArgLeuPheThrGluValLeuAspHisGlnGlu 780  
 QY 2598 TGTCTTATCATCACAGGTACTGAGCAGTACAGTGAATACACAGGCTATCCGAAACATAC 2657  
 Db 781 CysLeuIleIleThrGlyThrGluGlnTyrSerGluTyrThrGlyTyrAlaGluThrTyr 800  
 QY 2658 CCGTGGCCCGGAGCATGAGACAGGACGAAAGGACGACTGGCAGGCGCCAGCAT 2717  
 Db 801 ArgTTPAlaArgSerHisGluAspArgSerGluArgAspArgTTPGlnArgArgThrThr 820  
 QY 2718 GAGATCGTCCGATGAGCCCTCCACTTACAGCGCTACTCGACAGTGTGTGCCCGAG 2777  
 Db 821 GluIleValAlaIleAspAlaLeuHisPheArgTyrLeuAspGlnPheValProGlu 840  
 QY 2778 AAGATCAGACGGGAGCTTAACAGGCTTACTGTGATTTCTTCTGCTCGAGTTCTTCA 2837  
 Db 841 LysIleArgArgGluLeuAsnLysAlaTyrCysGlyPheLeuArgProGlyValSerSer 860  
 QY 2838 GAGAACTGCTGTCATGCTGACAGAACTGAGGCTGTGCTGCTTGGGGTGATGCT 2897  
 Db 861 GluAsnLeuSerAlaValAlaThrGlyAsnTrpGlyCysGlyAlaPheGlyGlyAspAla 880  
 QY 2898 AGACTAAAGCCTTAATACAGATCCTGCGAGCTGTGTAGCTGAGGAGACGTGGTTTAT 2957  
 Db 881 ArgLeuLysAlaLeuIleGlnIleLeuAlaAlaValAlaGluArgAspValValTyr 900  
 QY 2958 TTCACCTTTGGGACTCAGAACTGATGAGAGACATTTACAGCATGATCATTCCTCAT 3017  
 Db 901 PheThrPheGlyAspSerGluLeuMetArgAspIleTyrSerMetHisThrPheLeuThr 920  
 QY 3018 GAGAGAACTGACTGTTGGAGAGTATATACCTGCTGCTAGATATACATGAGAA 3077  
 Db 921 GluArgLysLeuThrValGlyGluValTyrLysLeuLeuLeuArgTyrTyrAsnGluGlu 940  
 QY 3078 TCGAAGAACTGCTCCACCCCGGACGACATCAAGCTTTTATCATATACCATGCA 3137  
 Db 941 CysArgAsnCysSerThrProGlyProAspIleLysLeuTyrProPheIleTyrHisAla 960  
 QY 3138 GTTGAGTCTCTGTACAGACACCAACCCAGCCGGGACAAAGACGGGGGCC 3188  
 Db 961 ValGluSerCysThrGlnThrThrAsnGlnProGlyGlnArgThrGlyAla 977

RESULT 4  
 ABG72278  
 ID ABG72278 standard; protein; 977 AA.  
 AC ABG72278;  
 XX  
 DT 13-MAR-2003 (first entry)  
 DE Bovine poly (ADP-ribose) glycohydrolase (PARG) enzyme.  
 XX  
 KW Bovine; poly (ADP-ribose) glycohydrolase; PARG; PARG expression;  
 KW cellular response; DNA damage; neoplastic disorder inducing agent;  
 KW oxidative stress; neoplastic disorder; myocardial infarction;  
 KW vascular stroke; neurodegenerative disease; Alzheimer's disease;  
 KW Parkinson's disease; Huntington's disease; inborn genetic error;  
 KW reperfusion; ischaemia; aging; neurotoxicity; PARG activity; cytostatic;  
 KW neuroprotective; neotropic; antiparkinsonian; cardiac; vasotrophic;  
 KW anticonvulsant; cerebroprotective; enzyme.  
 XX  
 OS Bos taurus.  
 XX  
 XX US2002132328-A1.  
 XX  
 PD 19-SEP-2002.  
 XX  
 XX 09-OCT-2001; 2001US-00973451.  
 XX

PR 01-MAY-1998; 98US-0083768P.  
 PR 30-APR-1999; 99US-00302812.  
 XX (JACO/) JACOBSON M K.  
 PA (JACO/) JACOBSON E L.  
 PA (AMEJ/) AME J.  
 PA (LINW/) LIN W.  
 XX Jacobson MK, Jacobson EL, Ame J, Lin W;  
 PI WPI: 2003-155895/15.  
 XX N-PSDB; ABX14477.  
 DR New nucleic acid molecule encoding a polypeptide with poly(ADP-ribose)  
 PT glycohydrolase activity, for preventing, treating, or ameliorating a  
 PT disease condition, e.g. neoplastic disorder, myocardial infarction or  
 PT vascular stroke.  
 XX Claim 28; Fig 16; 86pp; English.  
 PS The present invention relates to the isolation of poly(ADP-ribose)  
 CC glycohydrolase (PARG) from several species, and the polynucleotide  
 CC sequences encoding them. Methods for inhibiting PARG expression or  
 CC overexpressing PARG are also disclosed. PARG is involved in the cellular  
 CC response to DNA damage, and is associated with the body's response to  
 CC neoplastic disorder inducing agents and oxidative stress. The  
 CC polynucleotide sequences encoding PARG and PARG modulators are useful for  
 CC preventing, treating, or ameliorating diseases such as neoplastic  
 CC disorders, myocardial infarction, vascular stroke, neurodegenerative  
 CC disorders, (e.g. Alzheimer's disease, Parkinson's disease or Huntington's  
 CC disease), inborn genetic errors, reperfusion following ischaemia, aging,  
 CC and neurotoxicity. The polynucleotide sequences are also useful in gene  
 CC therapy. The methods are useful for identifying an agent that modulates  
 CC PARG activity, identifying a mutant PARG allele in an individual, or  
 CC screening candidate molecules for PARG modulating activity. The present  
 CC sequence represents bovine PARG enzyme  
 XX Sequence 977 AA;  
 SQ  
 Alignment Scores:  
 Pred. No.: 0 Length: 977  
 Score: 5184.00 Matches: 977  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 71.58% Indels: 0  
 DB: Gaps: 6  
 US-09-302-812-1 (1-4070) x ABG72278 (1-977)  
 QY 258 ATGAGTGGGGCCCCCGGCTGTGAGCCCTGCACCAAGGACCCCGCTGGAGCGCGTGA 317  
 Db 1 MetSerAlaGlyProGlyCysGluProCysThrLysArgProArgTrpAspAlaAla 20  
 QY 318 ACTTCTCCGCGCGCTCGGACGCGCGAGCTTCCCGGACGAGGAGCGCGCTCTC 377  
 Db 21 ThrSerProAlaAlaSerAspAlaArgSerPheProGlyArgGlnArgValLeu 40  
 QY 378 GATTCCAGGACGCTCCGGTGCAGTTTCAGGTCCTCCGCTCCGCTCAGCTCGCCCTG 437  
 Db 41 AspSerLysAspAlaProValGlnPheArgValProProSerSerSerGlyCysAlaLeu 60  
 QY 438 GCGCGCGGGGACAGCACCGGACGCGGACCTCTCTGTTTTCACACAGAGACTATA 497  
 Db 61 GlyArgAlaGlyGlnHisArgGlySerAlaThrSerLeuValPheLysGlnLysThrIle 80  
 QY 498 ACCAGTTGGATGACACTAAAGGAATCAACACAGTTGAATCAGAAAGTTTGCATAGTAA 557  
 Db 81 ThrSerTrpMetAspThrLysGlyIleLysThrValGluSerGluSerLeuHisSerLys 100  
 QY 558 GAAACACAAATACAGAGAGAAATCCATGATGATCTCTGACAAAAGAGTACTTTAT 617  
 Db 101 GluAsnAsnThrArgGluGluSerMetMetSerValGlnLysAspAsnPheTyr 120

QY 618 CAACATACATGGAAAAATTAGAAAAATGTTTCTCAGCTAGCTTTTGTATAGTCCACCAGTT 677  
 Db 121 GlnHisasnMetGluLysLeuGluasnValSerGlnLeuGlyPheAspLysSerProVal 140  
 QY 678 GAAAAAGGTACACAGTATTTGAAGCAGCATCAGACTCGGGCTATGCTGAAGTGCACAAAT 737  
 Db 141 GluLysGlyThrGlnTyrLeuLysGlnHisGlnThrAlaAlaMetCysLysTrpGlnAsn 160  
 QY 738 GAAGGGCCACACTCAGAACGGCTTTTGGAAAGTGAACCTCCAGCGGTAACTCTGGTACCA 797  
 Db 161 GluGlyProHisSerGluArgLeuLeuGluSerGluProAlaValThrLeuValPro 180  
 QY 798 GAGCAGATTCACTAATGCTTAATGCTCGATCAGTCGTCCCAAGAGATGATCACAGTGACACA 857  
 Db 181 GluGlnPheSerAsnAlaAsnValAspGlnSerProLysAspAspHisSerAspThr 200  
 QY 858 AATAGTAGAGGAGTAGAGATAATCAGCAGTTTTTGACATGTAAAGCTTGCGAATGCA 917  
 Db 201 AsnSerGluGluSerArgAspAsnGlnGlnPheLeuThrHisValLysLeuAlaAsnAla 220  
 QY 918 AAGCAGACGATGGAGATGAACAGGGCAGAGAGCCAGAACCCAGAAAGTGTGGCAAG 977  
 Db 221 LysGlnThrMetGluAspGluGlnGlyArgGluAlaArgSerHisGlnLysCysGlyLys 240  
 QY 978 GCTTCCCATCTCTSCAGAACCTGTCCAGGGTGTACAGAGGAGGACAGAGTGTGTGCC 1037  
 Db 241 AlaCysHisProAlaGluAlaCysAlaGlyCysGlnGlnGluThrAspValValSer 260  
 QY 1038 GAGAGCCCTTGTCCGACACTGCTCTGAGGATGTTGCTGAGTCTGAAATGCCAAC 1097  
 Db 261 GluSerProLeuSerAspThrGlySerGluAspValGlyThrGlyLeuLysAsnAlaAsn 280  
 QY 1098 AGATTGAATAGACAAGAAAGTAGTCTAGGAAATCTCTCCATTTGAGAAAGAAAGTGAA 1157  
 Db 281 ArgLeuAsnArgGlnGluSerSerLeuGlyAsnSerProProPheGluLysGluSerGlu 300  
 QY 1158 CTTGAGTCCACCAATGGATGATGATTAATCCAAAAATAGTTGTTCAGATTTCAGAACGAGT 1217  
 Db 301 ProGluSerProMetAspValAspAsnSerLysAsnSerCysGlnAspSerGluAlaAsp 320  
 QY 1218 GAAGAGACAGTCCAGGTTTTCATGACAGAGGAGATAGCAGTTCTCTCAAAACAGCAAT 1277  
 Db 321 GluGluThrSerProGlyPheAspGluGlnGluAspSerSerAlaGlnThrAlaAsn 340  
 QY 1278 AAACCTTCAAGGTTCCAAACAGAGAGCTGACACTGAGTTGAGGAAGCGCTCTGTCT 1337  
 Db 341 LysProSerArgPheGlnProArgGluAlaAspThrGluLeuArgLysArgSerSerAla 360  
 QY 1338 AAGGGAGGTGAGATTTCGATTTCCAAATTTGAAGGAGGAGAGATCGAGCTGGAATG 1397  
 Db 361 LysGlyGlyGluLeuArgLeuHisPheGlnPheGluGlyGlyGluSerArgAlaGlyMet 380  
 QY 1398 AATGATGTGAATGCCAAACCTGGAAGTACTTCTAGCTGAATGTAGAGTGCAGAAAT 1457  
 Db 381 AsnAspValAsnAlaLysArgProGlySerThrSerSerLeuAsnValGluCysArgAsn 400  
 QY 1458 TCTAAGCAAATGGGAGAAAGGATTTCTAAATCACAGATCATTTTCATGAGAGTGCCCAA 1517  
 Db 401 SerLysGlnHisGlyArgLysAspSerLysLeuThrAspHisPheMetArgValProLys 420  
 QY 1518 GCAGAGACAAAAGAAAGAACATGTGAATGAAATCAACATCAAGAACAGAAAGAGATC 1577  
 Db 421 AlaGluAspLysArgLysGluGlnCysGluMetLysHisGlnArgThrGluArgLysLeu 440  
 QY 1578 CCTAAATACATCCACTCTCCTCTCCAGATAAGAAATGGCTTGGAACTCTCTATTGAG 1637  
 Db 441 ProLysTyrLeuProProHisLeuSerProAspLysLysTrpLeuGlyThrProLysGlu 460  
 QY 1638 GAGATGAGGAGAAATGCCAAGGTGTGGATCCGGTCTCCCTTGGAGACCATCTGCCAAT 1697  
 Db 461 GluMetArgArgMetProArgCysGlyLeuArgLeuProLeuArgProSerAlaAsn 480  
 QY 1698 CACACAGTACTATTCGGGTAGATCTTTTGGATAGGAGAGTCTCTAAACCTTTCCCA 1757

Db 481 HisThrValThrIleArgValAspLeuLeuArgIleGlyGluValProLysProPhePro 500  
 QY 1758 ACACATTTTAAAGATTTTGGGCAACAACATGTTAAGATGCTTCTTTCAGAACAAAC 1817  
 Db 501 ThrHisPheLysAspLeuTrpAspAsnLysHisValLysMetProCysSerGluGlnAsn 520  
 QY 1818 TTGTACCTCTGGAAGATGAGATGAGTGGAGCGAGCTGCAGCGCCGCTGGAACTCAT 1877  
 Db 521 LeuTyrProValGluAspGluAsnGlyGluArgAlaAlaGlySerArgTrpGluLeuLe 540  
 QY 1878 CAGACTGCATCTTCAACAGGCTCACCTCGGCCCCAGAACCTCGAAGGATGCTATTCTGAAG 1937  
 Db 541 GlnThrAlaLeuLeuAsnArgLeuThrArgProGlnAsnLeuLysAspAlaIleLys 560  
 QY 1938 TACAATGTGCATATCTTAAGAAATGGGACTTTACAGCTTTGATGTTCTGGGATAAG 1997  
 Db 561 TyrAsnValAlaTyrSerLysLysTrpAspPheThrAlaLeuIleAspPheTrpAspLys 580  
 QY 1998 GTACTAGAAGACAGAGCTCAACACTTGTATCAGTCCATCTTTCCTGATATGCTGAAA 2057  
 Db 581 ValLeuGluGluAlaGluAlaGlnHisLeuTyrGlnSerIleLeuProAspMetValLys 600  
 QY 2058 ATTGCATCTCTGTGCCAAATATTTGTACCCAGGCAATACCACTCTCTGAAAACAGAGATG 2117  
 Db 601 IleAlaLeuCysLeuProAsnIleCysThrGlnProIleProLeuLeuLysGlnLysMet 620  
 QY 2118 AATCATTTCCATCACAATGTTCACAGAACAGATGCCAGTCTTTTAGCTAATGCTTCTTC 2177  
 Db 621 AsnHisSerIleThrMetSerGlnGluGlnIleAlaSerLeuLeuAlaAsnAlaPhePhe 640  
 QY 2178 TGCAGGTTTCCACAGCGCAATGCCAAGATGAAATCAGAGTATTCAGATTATCCAGATATT 2237  
 Db 641 CysThrPheProArgAsnAlaLysMetLysSerGluTyrSerSerTyrProAspIle 660  
 QY 2238 AACTTCAATCGGTTGTTTGAAGGACGTTTCATCAAGGAAACAGAGAACCTTAAACGCTC 2297  
 Db 661 AsnPheAsnArgLeuPheGluGlyArgSerSerArgLysProGluLysLeuLysThrLeu 680  
 QY 2298 TTCTGCTACTTAGAGAGTACAGAGAAAAACCCACTGGGTGCTGCACATTCAACAAGA 2357  
 Db 681 PheCysTyrPheArgArgValThrGluLysLysProThrGlyLeuValThrPheThrArg 700  
 QY 2358 CAGAGTCTTGAAGATTTTCCAGAGTGGGAAAGATGTGAAAACTCTCTGACTCGACTGAT 2417  
 Db 701 GlnSerLeuGluAspPheProGluTrpGluArgCysGluLysLeuLeuThrArgLeuHis 720  
 QY 2418 GTCACTTACGAAGGTACCATAGAGAAACCGGCCAGGGCATCTACAGTGCATTTTGA 2477  
 Db 721 ValThrTyrGluGlyThrIleGluGlyAsnGlyGlnGlyMetLeuGlnValAspPheAla 740  
 QY 2478 AACCGTTTGGTGGAGGTGGTGTAAACAGTGCAGGACTTGTGCAAGAAAGAAATCCGCTTT 2537  
 Db 741 AsnArgPheValGlyGlyGlyValThrSerAlaGlyLeuValGlnGluGluLeuArgPhe 760  
 QY 2538 TTAATCAACCCCTGAGTGTATTCTTTCACGGCTCTTCACTGAGGTGCTGATCATCATGA 2597  
 Db 761 LeuIleAsnProGluLeuIleValSerArgLeuPheThrGluValLeuAspHisAsnGlu 780  
 QY 2598 TGTCTTATCATCACAGGTACTGACGAGTACAGTGTATACACAGGCTATCCGAAACATAC 2657  
 Db 781 CysLeuIleIleThrGlyThrGluGlnTyrSerGluTyrThrGlyTyrAlaGluThrTyr 800  
 QY 2658 CGCTGGGCGCGAGCCATCAAGACAGAGCGGAAAGGAGCAGCTGGCAGAGGCGCAGCAT 2717  
 Db 801 ArgTrpAlaArgSerHisGluAspArgSerGluArgAspAspTrpGlnArgThrThr 820  
 QY 2718 GAGATCGTCGCCATCGACCGCTCTCCACTTCAGACGCTACTCTGCACAGTTTGTCCCGAG 2777  
 Db 821 GluIleValAlaIleAspAlaLeuHisPheArgArgTyrLeuAspGlnPheValProGlu 840  
 QY 2778 AGATCAGACGGGAGCTTAAAGGCTTACTGTGGATTTCTTCGCTCTCGAGTTTCTTCA 2837



841 LysIleArgArgGluLeuAsnLysAlaTyrCysGlyPheLeuArgProGlyValSerSer 860  
2838 GAGAACCTGCTGCGAGTGGCTACAGAACTGGGCTGTGGTGCCTTTGGGGGTGATGCT 2897  
861 GluAsnLeuSerAlaValAlaThrGlyAsnTrpGlyCysGlyAlaPheGlyAspAla 880  
2898 AGACTAAAGACCTTAATACAGATCTCGCAGCTGCTGAGTGTAGCTGAGCGAGACGTGTTAT 2957  
881 ArgLeuLysAlaLeuIleGlnIleLeuAlaAlaValAlaGluArgAspValValTyr 900  
2958 TTCACCTTTGGGACTCAGAACCTGATGAGACATTTACAGCATGATCATCTCTCACT 3017  
901 PheThrPheGlyAspSerGluLeuMetArgAspIleTyrSerMetHisThrPheLeuThr 920  
3018 GAGAGAAACTGACTGCTGGAGAGATATAGCTGCTGCTAGATATACATGAGAA 3077  
921 GluArgLysLeuThrValGlyGluValTyrLysLeuLeuLeuArgTyrTyrAsnGluGlu 940  
3078 TGCAGAACTGCTCCACCCCGGACGACATCAAGCTTTATCCATTCATATACCATGCA 3137  
941 CysArgAsnCysSerThrProGlyProAspIleLysLeuTyrProPheIleTyrHisAla 960  
3138 GTTGAGTCTGTATACAGACCAACACCCCGGACAAAGCGGGGCC 3188  
961 ValGluSerCysThrGlnThrThrAsnGlnProGlyGlnArgThrGlyAla 977

RESULT 5  
AAE25630  
ID AAE25630 standard; protein; 976 AA.  
AC AAE25630;  
XX  
DT 04-NOV-2002 (first entry)  
DE Human poly adenosine diphosphate-ribose glycohydrolase (PARG).  
KW Human; poly adenosine diphosphate-ribose glycohydrolase; PARG; enzyme;  
KW ADP-ribose; neoplastic disorder; genotoxic oxidative stress; neurotropic;  
KW cardiac disorder; neuronal disorder; reperfusion injury; neurotoxicity;  
KW Alzheimer's disease; Huntington's disease; Parkinson's disease; cardiant;  
KW cytosstatic; vasotrophic; neuroprotective; anticonvulsant; gene therapy;  
XX antisense therapy.  
OS Homo sapiens.  
XX  
XX USG395543-B1.  
XX  
XX 28-MAY-2002.  
XX  
XX 23-FEB-2000; 2000US-00511507.  
XX  
XX 01-MAY-1998; 98US-0083768P.  
XX  
XX 30-APR-1999; 99US-00302812.  
XX  
XX (KENT ) UNIV KENTUCKY RES FOUND.  
XX  
XX Jacobson MK, Jacobson EL, Ame J, Lin W;  
XX WPI; 2002-535641/57.  
XX  
XX N-PSDB; AAD42082.  
XX  
XX New nucleic acid molecule encoding bovine poly adenosine diphosphate-  
XX ribose glycohydrolase involved in cellular response to DNA damage,  
XX inhibition of which is useful for treating neoplastic disorders and  
XX neurodegenerative diseases.  
XX  
XX Claim 3; Col 55-60; 77pp; English.  
XX  
XX The invention relates to an isolated nucleic acid molecule which encodes  
XX a bovine poly (adenosine diphosphate (ADP)-ribose) glycohydrolase (PARG)  
XX which catalyses release of ADP-ribose from an ADP ribose polymer. The  
XX invention is useful as probes and primer molecules that can be used in  
XX hybridisation assays and polymerase chain reaction (PCR) amplification.

CC The knowledge of the nucleotide sequence of the PARG gene permits the  
CC preparation of antisense therapeutics containing sequences complementary  
CC to the mRNA of PARG gene. The antisense therapeutic is useful to treat  
CC neoplastic disorders and conditions caused by genotoxic oxidative stress  
CC e.g., cardiac disorders, neuronal disorders, reperfusion injury,  
CC neurotoxicity, Alzheimer's disease, Huntington's disease and Parkinson's  
CC disease. The invention is useful in gene therapy and antisense therapy.  
XX The present sequence is human PARG  
XX  
XX Sequence 976 AA;  
SQ  
Alignment Scores:  
Pred. No.: 0 Length: 976  
Score: 4635.50 Matches: 874  
Percent Similarity: 92.83% Conservative: 32  
Best Local Similarity: 89.55% Mismatches: 69  
Query Match: 64.01% Indels: 1  
DB: Gaps: 1  
US-09-302-812-1 (1-4070) x AAE25630 (1-976)  
QY 258 ATGAGTGGCGGCGGCTGTGAGCCCTGCACCAAGCGACCCGCTGGAGCGCGCTGCA 317  
DB 1 MetAsnAlaGlyProGlyCysGluProCysThrLysAlaThrArgTrpGlyAlaAlaThr 20  
QY 318 ACTTCTCCCGCGGCGCTCGGACGCGCGAGCTTCCCGGCGAGGAGCGCGGCGCTC 377  
DB 21 ThrSer---ProAlaAlaSerAspAlaArgSerPheProSerArgGlnArgValLeu 39  
QY 378 GATTCCAAAGGACGCTCCGCTGAGTTCAGGTCCTCCGCGCTCCCTCGTCAGCTCGCGCTG 437  
DB 40 AspProLysAspAlaHisValGlnPheArgValProProSerSerProAlaCysValPro 59  
QY 438 GCGCGGCGGCGAGCAGCAGCGCGCGCGCTCTCTGCTTTTCAACAGAGACTATA 497  
DB 60 GlyGlnAlaGlyGlnHisArgGlySerAlaThrSerLeuValPheLysGlnLysThrIle 79  
QY 498 ACCAGTTGGATGACACTAAAGAAATCAAGACAGTTCGAATCAGAAAGTTTCATAGTAAA 557  
DB 80 ThrSerTrpMetAspThrLysGlyIleLysThrAlaGluSerGluSerLeuAspSerLys 99  
QY 558 GAAACCAACATACAGAGAGAAATCCATGATGAGTCTGTACAAAAGATAACTTTAT 617  
DB 100 GluAsnAsnAsnThrArgIleGluSerMetSerValGlnLysAspAsnPheTyr 119  
QY 618 CAACATACATGGAATAATTAGAAATGTTTCTCAGCTAGCTTTTGTAGTACACAGT 677  
DB 120 GlnHisAsnValGluLysLeuValAsnValSerGlnLeuSerLeuAspLysSerLeuThr 139  
QY 678 GAAAGAGTACACAGTATTTGAAGCAGCATCAGACTGCGGCTATGTGTAGTGGCAGAA 737  
DB 140 GluLysSerThrGlnTyrLeuAsnGlnHisGlnThrAlaAlaMetCysLysTrpGlnAsn 159  
QY 738 GAAGGCGCAGCTCAGACGCGCTTTTGAAAAGTGAACCTCCAGCGGTAACTCTGTACCA 797  
DB 160 GluGlyLysHisThrGlnLysLeuLeuGluSerGluProGlnThrValThrLeuValPro 179  
QY 798 GAGCAGTTTCAGTAAATGCTAATGCTCAGTCAGTCTCCCAAGAGATGATCAGTGCACA 857  
DB 180 GluGlnPheSerAsnAlaAsnIleAspArgSerProGlnAsnAspAspHisSerAspThr 199  
QY 858 AATAGTGGAGGAGTAGATTAATCAGCAGTTTGTGACATGTAAAGCTTCGGAATGCA 917  
DB 200 AspSerGluGluAsnArgAspAsnGlnGlnPheLeuThrValLysLeuAlaAla 219  
QY 918 AAGCAGCAGCATGAGATGAACAGGCGCAGAGAGCAGAGCCACAGCAAGTGTGCAAG 977  
DB 220 LysGlnThrThrGluAspGlnHisAlaArgGluAlaLysSerHisGlnLysCysSerLys 239  
QY 978 GCTTGCATCTCTCGCAGAACCTGTGTGAGGGGTGTTCAGAGAGAGACAGAGCTGTGTC 1037  
DB 240 SerCysHisProGlyGluAspCysAlaSerCysGlnGlnAspGluIleAspValValPro 259



QY	1038	GAGAGCCCTTGTGACACTGGCTCTGAGGAGTGTGTTGCTACTGACTGAAAAATGCCAAC	1097
Db	260	LySerProLeuSerAspValGlySerGluAspValGlyThrGlySerLysAsnAspAsn	279
QY	1098	AGATTGAATAGACAAGAAAGTACTCTAGGAAATCTCTCTCATTTTGAAAGAAAGTGAA	1157
Db	280	LeuLeuileArgGlnGlnSerCysLeuGlyAsnSerProProPheGluLeuGlySerGlu	299
QY	1158	CCTGAGTCACCAATGGATGTAGATTAATCCAAAAATAGTTGTCCAGATTCAGACGAGAT	1217
Db	300	ProGluSerProMetAspValAspAsnSerLysAsnSerCysGlnAspSerGluAlaAsp	319
QY	1218	GAAGACAAGTCCAGTTTGTATGTAACAGGAGAGATAGCAGTTCTGCTCAACACGACAAAT	1277
Db	320	GluGluThrSerProGlyPheAspGluGlnGluAspGlySerSerSerGlnThrAlaAsn	339
QY	1278	AAACCTTCAAGTTCCAAACCAACAGAGAGCTGACACTGAGTTGAGGAAGCGTCTCTGCT	1337
Db	340	LysProSerArgPheGlnAlaArgAspAlaAspIleGluPheArgLysArgGlySerThr	359
QY	1338	AAGGAGGTGAGATTCGATTACATTTCCAAATTTGAAGAGGAGAGGTGCGAGCTGGAATG	1397
Db	360	LysGlyGlyGluValArgLeuHisPheGlnPheGluGlyGlyGluSerArgThrGlyMet	379
QY	1398	AATGATGTGAATGCAACACGACCTGGGAAGTACTTCTAGCTGAATGTAGAGTGCAGAAAT	1457
Db	380	AsnAspLeuAsnAlaLysLeuProGlyAsnIleSerSerLeuAsnValGluCysArgAsn	399
QY	1458	TCTAAGCAACATGGGAGAAAGGATTTCAAATCACAGATCATTTCTATGAGGTGCCAAA	1517
Db	400	SerLysGlnHisGlyLysAspSerLysIleThrAspHisLeuMetArgLeuProllys	419
QY	1518	GCAGAGGCAAAAGAAAAGACATGTGAATGAACATCAAAGAACAGAAAGGAGATC	1577
Db	420	AlaGluAspArgArgLysGluGlnTrpGluThrLysHisGlnArgThrGluArgLysIle	439
QY	1578	CCTAAATACATTTCCACCTCACCTTTCTCCAGATAAGAAATGGCTTGGAACTCTATTGAG	1637
Db	440	ProllysTyrrValProProHisLeuSerProAspLysLysTrpLeuGlyThrProlleGlu	459
QY	1638	GAGATGAGGAGAATGCCAAGGTGGGATCCGGCTGCTCCCTTGAGACCATCTGCCAAT	1697
Db	460	GluMetArgArgMetProArgCysGlyIleArgLeuProlleuArgProSerAlaAsn	479
QY	1698	CACACAGTCACATTTCCGGTGATCTTTTGGCAATAGAGAAGTTCTCTAAACCTTTCCCA	1757
Db	480	HisThrValThrIleArgValAspLeuLeuArgAlaGlyGluValProllysProPhePro	499
QY	1758	ACACATTTTAAAGATTTGTGGACAACAGCATGTTAAGATGCTTATGATTCGTAAG	1817
Db	500	ThrHisTyrrLysAspLeuTrpAspAsnLysHisValLysMetProCysSerGluGlnAsn	519
QY	1818	TTGTACCTCTGGAAGATGAGTAATCGTCAGCGAGCTGCAGACGCCGTGGGAACCTATT	1877
Db	520	LeuTyrrProValGluAspGluAsnGlyGluArgThrAlaGlySerArgTrpGluLeuIle	539
QY	1878	CAGACTGCACTTCTCAACAGGCTCACTCGGCCCCAGAACCTGAAAGGATGCTATTCTGAAG	1937
Db	540	GlnThrAlaLeuLeuAsnLysPheThrArgProGlnAsnLeuLysAspAlaIleLeuLys	559
QY	1938	TACAATGTGGCATATTCTTAAGAAATGGGACTTTACAGTTCTGATTTCTGGGATAG	1997
Db	560	TyrAsnValAlaTyrrSerLysLysTrpAspPheThrAlaLeuIleAspPheTrpAspLys	579
QY	1998	GTACTAGAAGACAGAGCTCAACACTTGATCATCAGTCCATCTGCTCGCTGATATGGTAAA	2057
Db	580	ValLeuGluGluAlaGluAlaGlnHisLeuTyrrGlnSerIleLeuProAspMetValLys	599
QY	2058	ATTGCACTCTGTGCGCAATATTTGTACCAGCCAAATACCACTCCTCGAAACAGAGATG	2117
Db	600	IleAlaLeuCysLeuProAsnIleCysThrGlnProIleProLeuLeuLysGlnLysMet	619
QY	2118	AATCATTTCCATCAATGTGCAGGACAGATTTGCCAGTCTTTAGCTTAATGCTTTCTTC	2177

620	AsnHisSerIleThrMetSerGlnGluGlnIleAlaSerLeuLeuAlaAsnAlaPhePhe	639
2178	TGCACGTTTCCACGACGCAATCCCAAGATGAAATCAGAGTATTCAGATTATCCAGATATT	2237
640	CysThrPheProArgAsnAlaLysMetLysSerGluTyrSerSerTyrProAspIle	659
2238	AACTTCAAATCGGTTGTTTGAAGGACGTTTCATCAAGGAAACACAGAGAGCTTAAACGCTC	2297
660	AsnPheAsnArgLeuPheGluGlyArgSerSerArgLysProGluLysLeuLysThrLeu	679
2298	TTCTGCTACTTTAGAAAGAGTACACAGAGAAACCCACTGGCTGGTGCATATTCACAGA	2357
680	PheCysTyrPheArgArgValThrGluLysLysProThrGlyLeuValThrPheThrArg	699
2358	CAGAGCTCTTGAAGATTTTCCAGAGTGGGAAGATGTGAAAAACTCTGACTCGACTGCAT	2417
700	GlnSerLeuGluAspPheProGluTrpGluArgCysGluLysProLeuThrArgLeuHis	719
2418	GTCACTTACGAAGGTACCATAGAGGAAACGGCCACGGGCATGCTACAGTGGATTTCGA	2477
720	ValThrTyrGluGlyThrIleGluGluAsnGlyGlnGlyMetLeuGlnValAspPheAla	739
2478	AACCGTTTCGTGGAGCTGGTGTAAACCTGCAGCTGCAGGACTTGTCAAGAGAAATCCGCTTT	2537
740	AsnArgPheValGlyGlyValThrSerAlaGlyLeuValGlnGluLysLeuArgPhe	759
2538	TTAATCAACCTCAGTTGATTGTTTCCAGCTCTTCCAGCTGCTTCACTGAGTGTGGATCACA	2597
760	LeuIleAsnProGluLeuIleIleSerArgLeuPheThrGluValLeuAspHisAsnGlu	779
2598	TGCTTTATCATCACAGTACTGACGACGTACAGTGAATACACAGGCTATCCGGAACATAC	2657
780	CysLeuIleIleThrGlyThrGluGlnTyrSerGluTyrThrGlyTyrAlaGluThrTyr	799
2658	CGCTGGCCCGGAGCCATGAAAGACAGAGCGAAGGACGACCTGGCAGAGCGCAGCACT	2717
800	ArgTrpSerArgSerHisGluAspGlySerGluArgAspAspCysGluArgArgCysThr	819
2718	GAGATCTCGCCATCGAGCCCTCCACTTCAGACGCTTACCTCGACCAAGTTTGTCCCGAG	2777
820	GluIleValAlaIleAspAlaLeuHisPheArgTyrLeuAspGlnPheValProGlu	839
2778	AAGATCAGACGGGAGCTTAAACAGGCTTACTGTGGATTCTTCTGCTCGAGTTCTTCA	2837
840	LysMetArgArgGluLeuAsnLysAlaTyrCysGlyPheLeuArgProGlyValSerSer	859
2838	GAGAACCTGCTCGAGTGGCTACAGGAACTGGGGCTGTGGTCTTTGGGGGTGATGCT	2897
860	GluAsnLeuSerAlaValAlaThrGlyAsnTrpGlyCysGlyAlaPheGlyGlyAspAla	879
2898	AGACTTAAAGCTTAAATACAGATCCTGGCAGCTGCTGTAGCTGAGCGAGACGTGGTTAT	2957
880	ArgLeuLysAlaLeuIleGlnIleLeuAlaAlaAlaGluArgAspValValTyr	899
2958	TTCACTTTGGGAGCTCAGAACTGATGAGACATTTACACATGCATACATTCCTCACT	3017
900	PheThrPheGlyAspSerGluLeuMetArgAspIleTyrSerMetHisIlePheLeuThr	919
3018	GAGAGAAACTGACTGTTGGAGAGTATATAGCTGTCTACAGATATTACAAATGAAGAA	3077
920	GluArgLysLeuThrValGlyAspValTyrLysLeuLeuLeuArgTyrTyrAsnGluGlu	939
3078	TGCAGAAACTGCTCCACCCCGGACACAGACATCAAGCTTTATCCATTCATATACCATGCA	3137
940	CysArgAsnCysSerThrProGlyProAspIleLysLeuTyrProPheIleTyrHisAla	959
3138	GTTCAGTCTCTGTACACAGACCAACCAACCGCGGACAAAGGACGGGG	3185
960	ValGluSerCysAlaGluThrAlaAspHisSerGlyGlnArgThrGly	975

21	ThrSer---	ProAlaAlaSerAspAlaArgSerPheProSerArgGlnArgGargValIleu	39
378	GATTCCAAAGGAGCTCCGGTGCAGTTTCAGGGTCCCGCGCTCTCTCGTCAGGCTGC	CCCGCTG	437
40	AspProLysAspAlaHisValGlnPheArgValProProSerSerProAlaCysValPro	59	
438	GCCCGGGGGGACAGACACGAGGACAGCGCACCTCTCTTGTTTTCACACAGAGACTATA	497	
60	GlyGlnAlaGlyGlnHisArgGlySerAlaThrSerLeuValPheLysGlnLysThrIle	79	
498	ACCAGTTCGATCGACACTAAAGCAATCAAGACAGTTCGAATCAGAAAGTTTGCATAGTAAA	557	
80	ThrSerTrpMetAspThrLysGlyIleLysThrAlaGluSerGluSerLeuAspSerLys	99	
558	GAATAACAATAACAGAGAAGAATCCATGATCAGTTCTGTACAAAAGATTAACCTTTTAT	617	
100	GluAsnAsnAsnThrArgIleGluSerMetSerSerValGlnLysAspAsnPheTyr	119	
618	CAACATAACATCGAAAAATTAGAAATGTTCTCAGCTAGGTTTTCATAAGTCACCAAGT	677	
120	GlnHisAsnValGlnLysLeuValAsnValSerGlnLeuSerLeuAspLysSerLeuThr	139	
678	GAATAAGGTACACAGATATTGTAAGCAGCATCAGACTGCGGCTATGTGTAAAGTGGCAGAAT	737	
140	GluLysSerThrGlnTyrLeuAsnGlnHisGlnThrAlaAlaMetCysLysTrpGlnAsn	159	
738	GAAGGGCCACTCAGAACCGCTTTTGGAAAGTGAACCTCCAGCGCTAACTCTGGTACCA	797	
160	GluGlyLysHisThrGluGlnLeuLeuGluSerGluProGlnThrValThrLeuValPro	179	
798	GAGCAGTTTCAGTAAATGCTAATGTCATGATCAGTCTGCCAAAGGATGATCACAGTGACACA	857	
180	GluGlnPheSerAsnAlaAsnIleAspArgSerProGlnAsnAspAspHisSerAspThr	199	
858	AATAGTGAGGAGAGTAGAGATAATCAGCAGTTTTTGCACATGTAAGCTTGGCAATGCA	917	
200	AspSerGluGluAsnArgAspAsnGlnGlnPheLeuThrThrValLysLeuAlaAsnAla	219	
918	ARCAGACGATCGAAGATGAACAGGGCAGAGAAGCCAGAACCCACAGAAAGTGTGCGAAG	977	
220	LysGlnThrThrGluAspGluHisAlaArgGluAlaLysSerHisGlnLysCysSerLys	239	
978	GTTTGCATCCTCGCAGAAGCCTGTGCAGGGTGTCCAGCAGAGACAGACAGCTGGTGTCC	1037	
240	SerCysHisProGlyGluAspCysAlaSerCysGlnGlnAspGluIleAspValValPro	259	
1038	GAGAGCCCTGTGCGACACTGCTCTGAGAGTCTTGTAAGTGCATGAAAAATGCCAAC	1097	
260	LysSerProLeuSerAspValGlySerGluAspValGlyThrGlySerLysAsnAspAsn	279	
1098	AGATTGAATAGACAAGAAAATAGTCTAGGAAATTCCTCCATTGTAAGAAGAAAGTGAA	1157	
280	LysLeuIleArgGlnGluSerCysLeuGlyAsnSerProProPheGluLysGluSerGlu	299	
1158	CCTGAGTCACCAATGGATGTAGATAATTCCAAATAATGTTGTCAGGATTCAGAAAGCAGAT	1217	
300	ProGluSerProMetAspValaAspAsnSerLysAsnSerCysGlnAspSerGluAlaAsp	319	
1218	GAAGACAGACTCCAGGTTTTGTATGAACAGCAAGATAGCAGTTCTGCTCAACAGCAAAT	1277	
320	GluGluThrSerProGlyPheAspGluGlnLysAspGlySerSerSerGlnThrAlaAsn	339	
1278	AAACCTTCAAGGTTCCAAACCAAGAGAGTCGACACTGAGTTGAGGAGCGGTCTCTGCT	1337	
340	LysProSerArgPheGlnAlaArgAspAlaAspIleGluPheArgLysArgTyrSerThr	359	
1338	AGGGAGGTGAGATTCGATTACATTTCCAAATTTCAAGCAGACAGACTCGAGCTGGGAATG	1397	
360	LysGlyGlyGluValArgLeuHisPheGlnPheGluGlyGlyGluSerArgThrGlyMet	379	
1398	AATGATGTGAATGCCAAACGACCTGGGAAGTACTTCTAGCTGAATGTAGAGTGCAGAAAT	1457	
380	AsnAspLeuAsnAlaLysLeuProGlyAsnIleSerSerLeuAsnValGluCysArgAsn	399	

[illegible]

QY	2538	TTAATCAACCTGAGTTGATTTCTTCAACGGCCTTCACTGAGGTGCTGATCACAATGAA	2539
Db	760	LeuIleAsnProGluLeuIleIleSerArgLeuPheThrGluValLeuAspHisAsnGlu	779
QY	2598	TGTCCTTATCATCACAGTACTCAGACGATGACAGTGAATACACAGGTATCCGGAACATAC	2657
Db	780	CysLeuIleIleThrGlyThrGluGlnTyrSerGluTyrThrGlyTyrAlaGluThrTyr	799
QY	2658	CGCTGGGCGGAGCCATCAAGACAGGACGGAAGGAGGACGATGGCAGAGGCGCAGACT	2717
Db	800	ArgTyrSerArgSerHisGluAspGlySerGluArgAspAspCysGluArgArgCysThr	819
QY	2718	GACATCGTCGCCATCAGACGCCCTCCACTTTCAGACGCTACTCGACCATTTGTGCCGAG	2777
Db	820	GluIleValAlaIleAspAlaLeuHisPheArgArgTyrLeuAspGlnPheValProGlu	839
QY	2778	ARGATCAGACGGAGCTTAACAAGGCTTACTTGGATTTCTTCGCTCGAGTTTCTTCA	2837
Db	840	LysMetArgArgGluLeuAsnLysAlaTyrCysGlyPheLeuArgProGlyValSerSer	859
QY	2838	GAGAACCTGTCAGTGGCTACAGAAAACCTGGGGCTGTGGTGCCTTTGGGGGTGATCT	2897
Db	860	GluAsnLeuSerAlaValAlaThrGlyAsnTyrGlyCysGlyAlaPheGlyGlyAspAla	879
QY	2898	AGACTAAAGCCCTTAATACAGATCCTGGCAGCTGCCTAGCTCAGCAGACGTGGTTAT	2957
Db	880	ArgLeuLysAlaLeuIleGlnIleLeuAlaAlaAlaGluArgAspValValTyr	899
QY	2958	TTACCTTTGGGACTCAGAACTGATCAGACATTTACAGCATGCATACATTCCTCACT	3017
Db	900	PheThrPheGlyAspSerGluLeuMetArgAspIleTyrSerMethHisIlePheLeuThr	919
QY	3018	GAGAGGAAACTGACTGTTGGAGAAGTATATAAGCTGCTACGATATTCAATGAAGAA	3077
Db	920	GluArgLysLeuThrValGlyAspValTyrLysLeuLeuLeuArgTyrTyrAsnGluGlu	939
QY	3078	TGCAGAAACTGCTCCACCCCGGACACAGATCAAGCTTTATCCATTTCATATCCATGCA	3137
Db	940	CysArgAsnCysSerThrProGlyProAspIleLysLeuTyrProPheIleTyrHisAla	959
QY	3138	GTTTCAGTCTGTACACAGACCACCAACCGCCGGGACAAAGGACGGGG 3185	
Db	960	ValGluSerCysAlaGluThrAlaAspHisSerGlyGlnArgThrGly 975	
RESULT 7			
ID	AAU76012		
XX	AAU76012	standard; protein; 976 AA.	
AC	AAU76012;		
XX			
DT	08-MAY-2002	(first entry)	
XX			
DE	Human poly (ADP-ribose) glycohydrolase (PARG) protein sequence.		
XX			
KW	Human; poly(ADP-ribose) glycohydrolase; PARG; PARG modulator; ADP-ribose;		
KW	adenosine diphosphate-ribose; DNA repair; apoptosis; neoplasia; cancer;		
KW	inherited genetic disease; myocardial infarction; vascular stroke; aging;		
KW	neurodegeneration; Huntington's disease; Parkinson's disease;		
KW	Alzheimer's disease; neurotoxicity.		
XX			
OS	Homo sapiens.		
XX			
FN	US6333148-B1.		
XX			
PD	25-DEC-2001.		
XX			
PF	30-APR-1999;	99US-00302812.	
XX			
PR	01-MAY-1998;	98US-0083768P.	
XX			
PA	(KENT ) UNIV KENTUCKY RES FOUND.		
PT	Jacobson MK, Jacobson EL, Ame J, Lin W;		

XX WPI; 2002-153820/20.  
 DR N-PSDB; ABK14494.  
 XX  
 PT Screening compounds for modulation of poly(ADP-ribose) glycohydrolase,  
 useful potentially for treating diseases associated with DNA damage, e.g.  
 cancer.  
 PT  
 XX  
 PS  
 XX Claim 3; Col 55-60; 80pp; English.  
 XX  
 CC The present invention relates to a new method for screening compounds for  
 CC ability to modulate activity of an enzyme that hydrolyses ADP (adenosine  
 CC diphosphate)-ribose from an ADP-ribose polymer. The compounds are  
 CC inhibitors or activators of PARP (poly(ADP-ribose) glycohydrolase) and  
 CC are used to treat or prevent any condition associated with DNA damage,  
 CC e.g. neoplasia, inherited genetic diseases, myocardial infarction,  
 CC vascular stroke, aging and neurodegeneration e.g. Huntington's,  
 CC Parkinson's or Alzheimer's diseases, or neurotoxicity generally.  
 CC Compounds identified by the new method are more effective than known  
 CC inhibitors and have fewer side effects. The present amino acid sequence  
 CC represents the human PARP protein of the invention. This protein is one  
 CC of several PARP proteins (AAU75799 and AAU76012-AAU76015) of the  
 CC invention  
 XX  
 XX Sequence 976 AA;  
 SQ  
 Alignment Scores:  
 Pred. No.: 0 Length: 976  
 Score: 4635.50 Matches: 874  
 Percent Similarity: 92.83% Conservative: 32  
 Best Local Similarity: 89.55% Mismatches: 69  
 Query Match: 64.01% Indels: 1  
 DB: 5 Gaps: 1  
 US-09-302-812-1 (1-4070) x AAU76012 (1-976)  
 QY 258 ATGAGTGGGGCCCGCTGAGCTGAGCCGCTGACCAAGCGACCCGCTGGAGCGCGCTGCA 317  
 Db 1 MetAlaAlaGlyProGlyCysGluProCysThrLysAlaThrArgTrpGlyAlaAlaThr 20  
 QY 318 ACTTCTCCGCGCGCGCTCGAGCGCCGCGAGCTTCCCGCGAGCGAGAGCGCGCTCCTC 377  
 Db 21 ThrSer---ProAlaAlaSerAspAlaArgSerPheProSerArgGlnArgValLeu 39  
 QY 378 GATTCCAGAGCGCTCCGGTGCAGTTCAGGGTCCCGCGCTCCTCGTCAGGCTGGCGCCCTG 437  
 Db 40 AspProLysAspAlaHisValGlnPheArgValProProSerSerProAlaCysValPro 59  
 QY 438 GCGCGCGCGGACAGCAGCGCGAGCGCGCCACCTCTCTTGTGTTTCAACACAGAGACTATA 497  
 Db 60 GlyGlnAlaGlyGlnHisArgGlySerAlaThrSerLeuValPheLysGlnLysThrIle 79  
 QY 498 ACCAGTTGGATGGACACTAAAGGAATCAAGACAGTGAATGATGAGTGGTGGTGGTGGTGGT 557  
 Db 80 ThrSerTrpMetAspThrLysGlyLysLysThrAlaGluSerGluSerLeuAspSerLys 99  
 QY 558 GAAACACACATACAGAGAGATCCATGATGATGATGATGATGATGATGATGATGATGATGAT 617  
 Db 100 GluAsnAsnAsnThrArgGlyGluSerMetSerValGlnLysAspAsnPheTyr 119  
 QY 618 CAACATAACATGGAATAATAGAAAATGTTTCTCAGCTAGGTTTGTGATAGTCCACGATT 677  
 Db 120 GlnHisValGluLysLeuValAsnValSerGlnLeuSerLeuAspLysSerLeuThr 139  
 QY 678 GAAAAGGTACACAGTATTGAGCAGCATCAGACTGGGCTATGTGTAGTGGCAGAT 737  
 Db 140 GluLysSerThrGlnTyrLeuAsnGlnHisGlnThrAlaAlaMetCysLysTrpGlnAsn 159  
 QY 738 GAAGGCCACACTCAGACGGCTTTTGGAAAGTGAACCTCCAGCGCTAACTCTGGTACCA 797  
 Db 160 GluGlyLysHisThrGlnGlnLeuGluSerGluProGlnThrValThrLeuValPro 179  
 QY 798 GAGCAGTTTCAGTAATGCTAATGCTAGTCTGCTCCCAAGGATGATCAGTGCACACA 857

Db GluGlnPheSerAsnAlaAsnIleAspArgSerProGlnAsnAspAspHisSerAspThr 199  
 QY 858 AATAGTGGAGAGTAGAGTAATCAGCAGTCTTTGACACATGTAAAGCTTCGCAATGCA 917  
 Db 200 AspSerGluGluAsnArgAspAsnGlnGlnPheLeuThrValLysLeuAlaAsnAla 219  
 QY 918 AAGCAGCAGATGGAAGATGAACAGCGCAGAGAGCCAGAGCCAGAGCCAGAGTGTGGCAAG 977  
 Db 220 LysGlnThrThrGluAspGluHisAlaArgGluAlaLysSerHisGlnLysCysSerLys 239  
 QY 978 GCTTCCCATCTGCAAGACCTGTGTCAGGGTCTCAGAGGAGGAGGAGCAGACACCTGTGTCTC 1037  
 Db 240 SerCysHisProGlyGluAspCysAlaSerCysGlnGlnAspGluLysValValPro 259  
 QY 1038 GAGAGCCCTGTGTCGACACTGCTCTCAGGATGTTGGTACTGGAATGAAAATGCGCAAC 1097  
 Db 260 LysSerProLeuSerAspValGlySerGluAspValGlyThrGlySerLysAsnAspAsn 279  
 QY 1098 AGATTGAATAGACAAAGATAGTCTAGAAATTCCTCCATTTGAGAAAGAAAGTGAAG 1157  
 Db 280 LysLeuIleArgGlnGluSerCysLeuGlyAsnSerProPheGluLysGluSerGlu 299  
 QY 1158 CCTGAGTCACCAATCGATGTAGATAATCCAAAATAGTTGTCCAGATTCCAGAGCAGAT 1217  
 Db 300 ProGluSerProMetAspValAspAsnSerLysAsnSerCysGlnAspSerGluAlaAsp 319  
 QY 1218 GAAGAGACAAAGTCCAGGTTTTCATGAACAGAGAGATAGCAGTCTGCTCAACAGCAAT 1277  
 Db 320 GluGluThrSerProGlyPheAspGluGlnGluAspGlySerSerSerGlnThrAlaAsn 339  
 QY 1278 AAACCTTCAAGTTCCAACAGAGAGCTGACACTGAGTTGAGGAGCGCTCCTCTGCT 1337  
 Db 340 LysProSerArgPheGlnAlaArgAspAlaAspIleGluPheArgLysArgTyrSerThr 359  
 QY 1338 AAGGAGGTGAGATTTCGATTACATTTCCAAATTTGAAGAGGAGGAGAGTCTGAGTGGATG 1397  
 Db 360 LysGlyGlyGluValArgLeuHisPheGlnPheGlnGlyGlyGluSerArgThrGlyMet 379  
 QY 1398 AATGATGTGAATCCCAACAGACCTGGAAGTACTTCTAGCCTGAAATGATGAGTGGAGAAAT 1457  
 Db 380 AsnAspLeuAsnAlaLysLeuProGlyAsnIleSerSerLeuAsnValGluCysArgAsn 399  
 QY 1458 TCTAAGCAACATCGGAGAAAGGATTTCAAAATCAGACATCATTTTCATGAGAGTGGCCAAA 1517  
 Db 400 SerLysGlnHisGlyLysLysAspSerLysLysThrAspHisLeuMetArgLeuProLys 419  
 QY 1518 GCAGAGACAAAAGAAAAGCAATGTAATGAAATCAAAATCAAAAGCAAGAGAGAGATC 1577  
 Db 420 AlaGluAspArgArgLysGluGlnTrpGluThrLysHisGlnArgThrGluArgLysIle 439  
 QY 1578 CCTAATATCATTCACCTCCTCTTCCAGATAGAAATGGCTTGGTGGTGGTGGTGGTGGTGG 1637  
 Db 440 ProLysTyrValProProHisLeuSerProAspLysLysTrpLeuGlyThrProIleGlu 459  
 QY 1638 GAGATGAGGAGATGCCAAGGTGGGATCCGGCTGCTCCCTTGGAGACCATCTGCCAAT 1697  
 Db 460 GluMetArgArgMetProArgGlyLysGlyLeuLeuProLeuLeuArgProSerAlaAsn 479  
 QY 1698 CACACAGTACTATTTCGGGTAGATCTTTTGGCAATAGGAGAGTTCCTTAAACCTTTCCCA 1757  
 Db 480 HisThrValThrIleArgValAspLeuLeuArgAlaGlyGluValProLysProPhePro 499  
 QY 1758 ACACATTTTAAAGATTTGTGGGACAAACAGCATGTTTAAAGTGCCTTGTTCAGACAAAC 1817  
 Db 500 ThrHisTyrLysAspLeuTrpAspAsnLysHisValLysMetProCysSerGluGlnAsn 519  
 QY 1818 TTGTACCTCTGGAAGATGAATGAGTGTGAGAGCTGAGAGCTGAGAGCGAGCGGTGGGAATCT 1877  
 Db 520 LeuTyrProValGluAspGluAsnGlyGluArgThrAlaGlySerArgTrpGluLeuIle 539  
 QY 1878 CAGACTGCACTTCTCAACAGGCTCACTCGGCCCCAGAACCTGGAAGGATGCTATTCTGAAG 1937

Db 540 GlnThrAlaLeuLeuAsnLysPheThrArgProGlnAsnLeuLysAspAlaIleLeuLys 559  
QY 1938 TACAATGTGGCATATTCAGAAATGGGACTTTACAGCTTTGATGATTTCTGGGATAAG 1937  
Db 560 TyrAsnValAlaTyrSerLysLysTyrAspPheThrAlaLeuLysAspPheTrpAspLys 579  
QY 1998 GTACTAGAGAGAGAGAGCTCAACACTGTGATCAGTCCATCTTGTGCTGATATGCTGAAA 2057  
Db 580 ValLeuGluGluAlaGluAlaGlnHisLeuLysTyrGlnSerIleLeuProAspMetValLys 599  
QY 2058 ATTGCACCTCTGCTGCGCAATATTTGTATCCAGGCAATACACTCTCTGAAAACAGAGATG 2117  
Db 600 IleAlaLeuCysLeuProAsnIleCysThrGlnProIleProLeuLeuLysGlnLysMet 619  
QY 2118 AATCATTCATACATGTACAGAAACAGATGTCAGTCTTTAGCTAATGCTTCTTC 2177  
Db 620 AsnHisSerIleThrMetSerGlnGluGlnIleAlaSerLeuLeuAlaAsnAlaPhePhe 639  
QY 2178 TGCAGCTTTCCACGACGCAATGCCAAGATGAAATCAGAGTATTCAGTTATCCAGATATT 2237  
Db 640 CysThrPheProArgArgAsnAlaLysMetLysSerGlnTyrSerSerTyrProAspIle 659  
QY 2238 AACTTCATTCGTTGTTGAAGACGTTTCATCAAGAAACAGAGAACTTAAACGCTC 2297  
Db 660 AsnPheAsnArgLeuPheGluGlyArgSerArgLysProGluLysLeuLysThrLeu 679  
QY 2298 TTCTGCTACTTTAGAGAGTACAGAGAAACCACTGGTGGTGGTGCATTCACAAGA 2357  
Db 680 PheCysTyrPheArgArgValThrGluLysLysProThrGlyLeuValThrPheThrArg 699  
QY 2358 CAGAGTCTTGAAGATTTCCAGAGTGGGAAAGATGTGAAAACTCTCTGACTCGACTGCAT 2417  
Db 700 GlnSerLeuGluAspPheProGlnTyrGluArgCysGluLysProLeuThrArgLeuHis 719  
QY 2418 CTCACCTTACGAAAGTACCATAGAGAAACGCCAGGCGATGCTACAGTGGATTTGCA 2477  
Db 720 ValThrTyrGluGlyThrIleGluGluAsnGlyGlnGlyMetLeuGlnValAspPheAla 739  
QY 2478 AACCGTTTCGTTGGAGTGGTGTATACCACTGAGGACTGTGCGAAGAAATCCGCTTT 2537  
Db 740 AsnArgPheValGlyGlyValThrSerAlaGlyLeuValGlnGluLeuArgPhe 759  
QY 2538 TTAATCAACCTTGAGTTGTTTTCACGGCTCTTCACAGTGTGCTGATCACAATGAA 2597  
Db 760 LeuIleAsnProGluLeuIleSerArgLeuPheThrGluValLeuAspHisAsnGlu 779  
QY 2598 TGTCTTATCATCACAGGTACTGACGAGTACAGTGAATACACAGGCTATCCGAAACATAC 2657  
Db 780 CysLeuIleIleThrGlyThrGluGlnTyrSerGlnTyrThrGlyTyrAlaGluThrTyr 799  
QY 2658 CGCTGGCGCGGAGCCATGACAGACAGCGAGGAGGAGCTGCGCAGAGGCGCAGCT 2717  
Db 800 ArgTrpSerArgSerHisGluAspGlySerGluArgAspCysGluArgArgCysThr 819  
QY 2718 GAGATCGTCGCCATCGACGCCCTCCACTTCAGACGCTACTCGACAGTTGTGCCGAG 2777  
Db 820 GluIleValAlaIleAspAlaLeuHisPheArgTyrArgTyrLeuAspGlnPheValProGlu 839  
QY 2778 AGATCAGACGGAGCTTAAACAGGCTTACTGTGGATTTCTGCTCGTGGATTTCTTCA 2837  
Db 840 LysMetArgArgGluLeuAsnLysAlaTyrCysGlyPheLeuArgProGlyValSerSer 859  
QY 2838 GAGACCTGTCTCAGTGGCTACAGAACTCGGGCTGTGCTGCTTTGGGGGTGATGCT 2897  
Db 860 GluAsnLeuSerAlaValAlaThrGlyAsnTrpGlyCysGlyAlaPheGlyGlyAspAla 879  
QY 2898 AGACTAAACCTTATACAGATCTCGCAGCTGCTGTAGCTGAGCGAGACGTGGTTAT 2957  
Db 880 ArgLeuLysAlaLeuIleGlnIleLeuAlaAlaAlaAlaGluArgAspValValTyr 899  
QY 2958 TTCACCTTTGGGACTCAGAACTGTATGAGACATTTTACAGCATGCATATCTTCCTCACT 3017  
Db 900 PheThrPheGlyAspSerGluLeuMetArgAspIleTyrSerMetHisIlePheLeuThr 919

QY 3018 GAGAGGAAACTGACTGTGTGAGAGATATTAAGTCTGCTACGATATTCAATGAGAA 3077  
Db 920 GluArgLysLeuThrValGlyAspValTyrLysLeuLeuLeuArgTyrTyrAsnGluGlu 939  
QY 3078 TGCAGAAACTGCTCCACCCCGGACGACATCAGCTTTATCCATTCATATACCATGCA 3137  
Db 940 CysArgAsnCysSerThrProGlyProAspIleLysLeuTyrProPheIleTyrHisAla 959  
QY 3138 GTTGAGTCTCTGTACACAGACCAACACGCGGACAAAGAGCGGG 3185  
Db 960 ValGluSerCysAlaGluThrAlaAspHisSerGlyGlnArgThrGly 975  
RESULT 8  
ABG72279  
ID ABG72279 standard; protein; 976 AA.  
XX  
XX  
AC ABG72279;  
XX  
XX  
DT 13-MAR-2003 (first entry)  
XX  
XX  
DE Human poly(ADP-ribose) glycohydrolase (PARG) enzyme.  
XX  
XX  
KW Human; poly(ADP-ribose) glycohydrolase; PARG; PARG expression;  
KW cellular response; DNA damage; neoplastic disorder inducing agent;  
KW oxidative stress; neoplastic disorder; myocardial infarction;  
KW vascular stroke; neurodegenerative disorder; Alzheimer's disease;  
KW Parkinson's disease; Huntington's disease; inborn genetic error;  
KW reperfusion; ischaemia; aging; neurotoxicity; PARG activity; cytosolic;  
KW neuroprotective; neotropic; antiparkinsonian; cardiant; vasotropic;  
KW anticonvulsant; cerebroprotective; enzyme.  
XX  
XX  
OS Homo sapiens.  
XX  
XX  
PN US2002132328-A1.  
XX  
XX  
PD 19-SEP-2002.  
XX  
XX  
PF 09-OCT-2001; 2001US-00973451.  
XX  
XX  
PR 01-MAY-1998; 98US-0083768P.  
PR 30-APR-1999; 99US-00302812.  
XX  
XX  
XX (JACO/) JACOBSON M K.  
PA (JACO/) JACOBSON E L.  
PA (AMEJ/) AME J.  
PA (LINW/) LIN W.  
XX  
XX  
PI Jacobson MK, Jacobson EL, Ame J, Lin W;  
XX  
XX  
DR WPI; 2003-155895/15.  
DR N-PSDB; ABX14478.  
XX  
XX  
PT New nucleic acid molecule encoding a polypeptide with poly(ADP-ribose)  
PT glycohydrolase activity, for preventing, treating, or ameliorating a  
PT disease condition, e.g. neoplastic disorder, myocardial infarction or  
PT vascular stroke.  
XX  
XX  
PS Claim 28; Fig 16; 86pp; English.  
XX  
XX  
CC The present invention relates to the isolation of poly(ADP-ribose)  
CC glycohydrolase (PARG) from several species, and the polynucleotide  
CC sequences encoding them. Methods for inhibiting PARG expression or  
CC overexpressing PARG are also disclosed. PARG is involved in the cellular  
CC response to DNA damage, and is associated with the body's response to  
CC neoplastic disorder inducing agents and oxidative stress. The  
CC polynucleotide sequences encoding PARG and PARG modulators are useful for  
CC preventing, treating, or ameliorating diseases such as neoplastic  
CC disorders, myocardial infarction, vascular stroke, neurodegenerative  
CC disorders (e.g. Alzheimer's disease, Parkinson's disease or Huntington's  
CC disease), inborn genetic errors, reperfusion following ischaemia, aging,  
CC and neurotoxicity. The polynucleotide sequences are also useful in gene  
CC therapy. The methods are useful for identifying an agent that modulates

CC PARG activity, identifying a mutant PARG allele in an individual, or  
 CC screening candidate molecules for PARG modulating activity. The present  
 CC sequence represents human PARG enzyme

XX  
 SQ Sequence 976 AA;

Alignment Scores:  
 Pred. No.: 0 Length: 976  
 Score: 4635.50 Matches: 874  
 Percent Similarity: 92.83% Conservative: 32  
 Best Local Similarity: 89.55% Mismatches: 69  
 Query Match: 64.01% Indels: 1  
 DB: 6 Gaps: 1

US-09-302-812-1 (1-4070) x ABG72279 (1-976)

QY	258	ATGAGTGGCGGCGCGCTGTGAGCCCTGCACCAAGCGACCCCGCTGGGACGCGCTGCA	317
DB	1	MetAsnAlaGlyProGlyCysGluProCysThrLysAlaThrArgTrpGlyAlaAlaThr	20
QY	318	ACTTCTCCGCGCGCTCGGACGCGCGGAGCTTCCCGGCGAGCGGCGCGCTCTC	377
DB	21	ThrSer---ProAlaAlaSerAspAlaArgSerPheProSerArgGlnArgValLeu	39
QY	378	GATTCACAGGACCTCCGGTGCAGTTCCAGGTCCTCCGCTCGCTCAGGCTCGCCCTG	437
DB	40	AspProLysAspAlaHisValGlnPheArgValProProSerSerProAlaCysValPro	59
QY	438	GGCGGCGCGGACAGCACCGGAGCGCGCCACCTCTCTGTTTCAAAAGAGACTATA	497
DB	60	GlyGlnAlaGlyGlnHisArgGlySerAlaThrSerLeuValPheLysGlnThrIle	79
QY	498	ACCAGTTGGATGACACTAAAGGATCAAGACAGTTGAAATCAGAAATTTGCATGATA	557
DB	80	ThrSerTrpMetAspThrLysGlyIleLysThrAlaGluSerGluSerLeuAspSerLys	99
QY	558	GAACAACACATACAGAGAAGATCCATGATGATTCGTACAAAAGATAACTTTTAT	617
DB	100	GluAsnAsnThrArgIleGluSerMetSerSerValGlnLysAspAsnPheTyr	119
QY	618	CAACATAACATGAAAATAATAGAAAATGTTTCTCAGCTAGGTTTGTGATAAGTCACCA	677
DB	120	GlnHisAsnValGluLysLeuValAsnValSerGlnLeuSerLeuAspLysSerLeuThr	139
QY	678	GAAGAAGGTACACAGTATTTGAAGCACATCAGCTCGGCTATGCTAGTGTGACAGAT	737
DB	140	GluLysSerThrGlnTyrLeuAsnGlnHisGlnThrAlaAlaMetCysLysTrpGlnAsn	159
QY	738	GAAGGCCACACTCAGACGCGCTTTTGGAAAGTGAACCTCCAGCGGTAACTCTGGTACCA	797
DB	160	GluGlyLysHisThrGluGlnLeuLeuGluSerGluProGlnThrValThrLeuValPro	179
QY	798	GAGCAGTTTCAGTAAATGCTGATCGATCAGTCGTCGCCCAAGGATGATCAGTGCACA	857
DB	180	GluGlnPheSerAsnAlaAsnIleAspArgSerProGlnAsnAspAspHisSerAspThr	199
QY	858	AATAGTCAGGAGTACAGATAATCAGCAGTTTTTGGACATGTAAGCTTGCGAATGCA	917
DB	200	AspSerGluGluAsnArgAspAsnGlnPheLeuThrThrValLysLeuAlaAsnAla	219
QY	918	AAGCAGACGATGGAAGATGAACAGGCGCAGAGAAGCCAGACCCAGAGTGTGGCAAG	977
DB	220	LysGlnThrThrGluAspGluHisAlaArgGluAlaLysSerHisGlnLysCysSerLys	239
QY	978	CTTTGCCATCTCGCAGAGCCCTGTGAGGGTGTGACAGAGGAGGACAGAGCTGGTGCC	1037
DB	240	SerCysHisProGlyGluAspCysAlaSerCysGlnGlnAspGluIleAspValPro	259
QY	1038	GAGAGCCCTTGTGCGACACTGGCTCTCAGAGATGTTGGTACTGGAAGTGAAGTCCCAAC	1097
DB	260	LysSerProLeuSerAspValGlySerGluAspValGlyThrGlySerLysAsnAspAsn	279
QY	1098	AGATTGATACACAGAAAGTAGTCTAGGAAATTTCTCTCCATTTGAGAAAGAAAGTGAA	1157

DB	280	LysLeuIleArgGlnGluSerCysLeuGlyAsnSerProPheGluLysGluSerGlu	299
QY	1158	CTTGAGTCACCAATGATGATGATATATCCAAAATAGTTGTGAGATTCACAAAGCAGAT	1217
DB	300	ProGluSerProMetAspValAspAsnSerLysAsnSerCysGlnAspSerGluAlaAsp	319
QY	1218	GAAGAGACAAAGTCCAGGTTTTCATGAACAGAGAGATAGCAGTTCTGCTCAACAGCAAT	1277
DB	320	GluGluThrSerProGlyPheAspGluGlnGluAspGlySerSerSerGlnThrAlaAsn	339
QY	1278	AAACCTTCAAGTTCCAAACAGAGAGAGCTGACACAGTGTGAGAGAGGCTCTCTGCT	1337
DB	340	LysProSerArgPheGlnAlaArgAspAlaAspIleGluPheArgLysArgTyrSerThr	359
QY	1338	AAGGAGGCTGAGATTCGATTACATTTCCAAATTTGAAGGAGGAGAGAGTCCAGTGGAAATG	1397
DB	360	LysGlyGlyGluValArgLeuHisPheGlnPheGluGlyGlyGluSerArgThrGlyMet	379
QY	1398	AATGATGTCAATGCCAAACGACCTGGAAGTACTTCTAGCTGAATGATGAGTGCAGAAAT	1457
DB	380	AsnAspLeuAsnAlaLysLeuProGlyAsnIleSerSerLeuAsnValGluCysArgAsn	399
QY	1458	TCTAAGCAATGGGAGAAAGATTTCTAAATCAGACATCATTTTCATGAGTGCACCAAA	1517
DB	400	SerLysGlnHisGlyLysAspSerLysIleThrAspHisLeuMetArgLeuProLys	419
QY	1518	CGAGGACGACAAAGAAAGAAACAAATGTGAATGAAACATCAAGAAACAGAAAGAAATC	1577
DB	420	AlaGluAspArgArgLysGluGlnTrpGluThrLysHisGlnArgThrGluArgLysIle	439
QY	1578	CCTAAATACATTCACCTTCTCCAGATTAAGAAATGGCTTGGAACTCTTATTTAG	1637
DB	440	ProLysTyrValProProHisLeuSerProAspLysLysTrpLeuGlyThrProIleGlu	459
QY	1638	GAGATGAGAGAAAGCCAAAGTGTGGATCCGCTCCCTTGGACCATCTGCCAAT	1697
DB	460	GluMetArgArgMetProArgCysGlyIleArgLeuProLeuLeuLeuArgProSerAlaAsn	479
QY	1698	CACACAGTCACTATTCCGGTAGATCTTTTGGCAATAGGAGAGTTCCTAAACCTTTCCCA	1757
DB	480	HisThrValThrIleArgValAspLeuLeuArgAlaGlyGluValProLysProPhePro	499
QY	1758	ACACATTTTAAAGATTGTGGGACAAACAGCATGTTAAGATGCTTGTTCAGACACAAAC	1817
DB	500	ThrHisTyrLysAspLeuTrpAspAsnLysHisValLysMetProCysSerGluGlnAsn	519
QY	1818	TTGTACCTCTGTGAAGATGAGATGTTGAGCGAGCTGCAGCGAGCGGTGGGAATCATT	1877
DB	520	LeuTyrProValGluAspGluAsnGlyGluArgThrAlaGlySerArgTrpGluLeuIle	539
QY	1878	CAGACTGCACCTCTCAACAGGCTCACTCGGCCCCAGAACCTGAAGAGTATTTCTGAAG	1937
DB	540	GlnThrAlaLeuLeuAsnLysPheThrArgProGlnAsnLeuLysAspAlaIleLeuLys	559
QY	1938	TACATGTGCAATTTCTAAGAAATGGACTTTACAGCTTTGATTTGATTTCTGGGTAAG	1997
DB	560	TyrAsnValAlaTyrSerLysLysTrpAspPheThrAlaLeuIleAspPheTrpAspLys	579
QY	1998	GTACTAAGAACAGCAGAGCTCAACACTTCTATCAGTCCATCTTTCCTGATATGCTGAAA	2057
DB	580	ValLeuGluGluAlaGluAlaGlnHisLeuTyrGlnSerIleLeuProAspMetValLys	599
QY	2058	ATTGCACTGTCTGCCAAATATTTGTACCCAGCAATACACCTCTCTGAAACAGAAATG	2117
DB	600	IleAlaLeuCysLeuProAsnIleCysThrGlnProIleProLeuLeuLysGlnLysMet	619
QY	2118	AATCATTTCCATCACAATGTACAGGAACAGATTGCCAGTCTTTTAGCTAATGCTTTCTTC	2177
DB	620	AsnHisSerIleThrMetSerGlnGluGlnIleAlaSerLeuLeuAlaAsnAlaPhePhe	639
QY	2178	TGCAGCTTTCCACGACCAATGCCAAGTAAATCAGAGTATTTCCAGTATTTCCAGATATT	2237





438 GGCGCGGGGAGCAGACCGAGCGCGCCACCTCTCTGTTTTCACAAAGACATATA 497  
 60 GlyArgAlaGlyProHisArgGlyAsnAlaThrSerPheValPheLysGlnYsThrIle 79  
 498 ACCAGTTGGATGACACTAAAGAAATCAGACAGTTCGAATCAGAAAGTTTGCATAGTAA 557  
 80 ThrThrTrpMetAspThrLysGlyProLysThrAlaGluSerGlu-----SerLys 96  
 558 GAAACAAACAAATCAGAGAAAGATCCATGATGATCTCTGACAAAAGATAACTTTTAT 617  
 97 GluAsnAsnThrArgIleAspSerMetLysSerValGlnLysAspAsnPheTyr 116  
 618 CAACATAACATGAAAAATAGAAAATGTTTCTCAGCTAGCTGTTTGTATAAGTACACAGTT 677  
 117 ProHisLysValGluLysLeuGluAsnValProGlnLeuAsnLeuAspLysSerProThr 136  
 678 GAAAAAGGTACAGATATTGAGCAGCATCAGACTCGGCTATGTTGTAAGTGGCAGAAAT 737  
 137 GluLysSerSerGlnTyrLeuAsnGlnGlnThrAlaSerValCysLysTrpGlnAsn 156  
 738 GAAGGCGCACCTCAGAACGCGCTTTTGGAAAGTGAACCTCCAGCGGTAACCTCTGTACCA 797  
 157 GluGlyLysHisAlaGluLeuLeuAlaSerGluProAlaGlyThrProLeuPro 176  
 798 GAGCAGTTTCAGTAATGCTAATGTCGATCAGTCGTCGCCCAAGAGATGATCAGATGACACA 857  
 177 LysGlnLeuSerAsnAlaAsnIleGlnSerProHisThrAspAspHisSerAspThr 196  
 858 AATAGTAGGAGATGAGATTAATCAGACAGTTTGTGACATGTAAGCTTGCAGATGCA 917  
 197 AspHisGluGluAspArgAspAsnGlnGlnPheLeuThrProIleLysLeuAlaAsnThr 216  
 918 AAGCAGACGATGAGAGTGAACAGCGGCAGAGAGCCAGAGCCAGAGAGTGTGGCAAG 977  
 217 LysProThrValGlyAspGlyGln-----AlaArgSerAsnCysLysSerGly 233  
 978 GCTTGCCATCCTCGAAGACCTGTGCGAGGTTCAGCAGGAGGAGACAGAGTGTGTCC 1037  
 234 SerArgGlnSerValLysAspCysThrGlyCysGlnGlnGluValAspValLeuPro 253  
 1038 GAGACCCCTGTGCGCACTGGCTCTGAGGATGTTGTTACTGGAATGAAAATGCCAAC 1097  
 254 GluSerProLeuSerAspValGlyAlaGluAspIleGlyThrGlyProLysAsnAspAsn 273  
 1098 AGATTGAATAGACAAAGATAGTCTAGGAAATCTCTCCATTTGAGAAAGAAAGTGA 1157  
 274 LysLeuThrGlyGlnGluSerSerLeuGlyAspSerProPheGluLysGluSerGlu 293  
 1158 CTGAGTCAACATGATGATGATTAATTCAAAATAGTTGTGAGATTTCAGAGCAGAT 1217  
 294 ProGluSerProMetAspValAspAsnSerArgAsnSerCysGlnAspSerGluAlaAsp 313  
 1218 GAAGGACAGTCCAGTTTTCATGAAAGAGAGATAGCAGATTTCTCTCAAAACAGCAAT 1277  
 314 GluGluThrSerProValPheAspGluGlnAspAspArgSerSer---GlnThrAlaAsn 332  
 1278 AAACCTTCAAGGTTCCAAACCAAGAGAGCTGACATCTGATGAGGAAGCGGCTCTGCT 1337  
 333 LysLeuSerSerCysGlnAlaArgGluAlaAspGlyAspLeuArgLysArgTyrLeuThr 352  
 1338 AAGGAGGTGAGATTCGATTAATTTCCATTTGAGAGGAGAGAGTCCGAGCTGGAATG 1397  
 353 LysGlySerGluValArgLeuHisPheGlnPheGlu---GlyGluAsnAlaGlyThr 371  
 1398 AATGATGTGATGCCAACACACCTCGAAGTACTTCTAGCTGATGATGATGATGATGAT 1457  
 372 SerAspLeuAsnAlaLysProSerGlyAsnSerSerSerLeuAsnValGluCysArgSer 391  
 1458 TCTAAGCAACATCGGAGAGAGTTCCTAAATATCAGATCATTCATTCAGAGTGGCCAAA 1517  
 392 SerLysGlnHisGlyLysArgAspSerLysIleThrAspHisPheMetArgLysSerLys 411  
 1518 GCAGAGGACAAAAGAAAAGAAACAATGTGAAATGAAACATCAAGACAGAAAGAGATC 1577

412 SerGluAspArgArgLysGluGlnCysGluValArgHisGlnArgThrGluArgLysIle 431  
 1578 CCTAAATACATTCACCTCACCTCTCCAGATAAGAAATGGCTTGGAACTCCCTATTGAG 1637  
 432 ProLysTyrIleProProAsnLeuProProGluLysTrpLeuGlyThrProIleGlu 451  
 1638 GAGATGAGGAGATGCAAGAGTGTGGATCGGCTCGCTCCCTTGCAGACCATCTGCCAAT 1697  
 452 GluMetArgLysMetProArgCysGlyIleHisLeuProSerLeuArgProSerAlaSer 471  
 1698 CACACAGTGTGATTTGGGTAGATCTTTTCGAATAGGAGAGTTCCTAAACCTTCCCA 1757  
 472 HisThrValThrValArgValAspLeuLeuArgAlaGlyGluValProLysProPhePro 491  
 1758 ACACATTTTAAAGATTGTGGGACACACAGCATGTTAAGATGCTTCTCAGAACAAAC 1817  
 492 ThrHisTyrLysAspLeuTrpAspAsnLysHisValLysMetProCysSerGluGlnAsn 511  
 1818 TTGTACCTCTGTGAGATGAGATGTGAGCGAGCTGCAGCGACCGGTGGAACTCAT 1877  
 512 LeuTyrProValGluAspGluAsnGlyGluArgThrAlaGlySerArgTrpGluLeuIle 531  
 1878 CAGACTGCATCTTCAACAGGCTCCTCGGCCAGAACCTGAAGAGTCTTATTTCTGAG 1937  
 532 GlnThrAlaLeuLeuAsnLysPheThrArgProGlnAsnLeuLysAspAlaIleLeuLys 551  
 1938 TACATGTGGCATATTCTAAGAAATGGACTTTACAGCTTTGATTGATTTCTGGGATAAG 1997  
 552 TyrAsnValAlaTyrSerLysLysLysPheThrAlaLeuValAspPheTrpAspLys 571  
 1998 GTACTGAGAGAGCAGAGCTCAACACTTGTATCAGTCCATCTTGCCTTGATATGGTGAAA 2057  
 572 ValLeuGluGluAlaGluAlaGlnHisLeuTyrGlnSerIleLeuProAspMetValLys 591  
 2058 ATTGCATCTCTCTGCCAAATATTGTTACCCAGCCCATACCACTCTCTGAACAGAGATG 2117  
 592 IleAlaLeuCysLeuProAsnIleCysThrGlnProIleProLeuLeuLysGlnLysMet 611  
 2118 AATCATTCATCACTACATGCTCAGAGACAGATGCGAGTCTTTAGCTTAATCTTTCTTC 2177  
 612 AsnHisSerValThrMetSerGlnGlnIleAlaSerLeuLeuAlaAsnAlaPhePhe 631  
 2178 TGCAGTTTCCACGACGCAATGCCAAGATGAAATCAGAGTATTCAGTTATCCAGATATT 2237  
 632 CysThrPheProArgArgAsnAlaLysMetLysSerGluTyrSerTyrProAspIle 651  
 2238 AACTCAATCGGTTGTTTGAAGAGCTTCATCAAGAAACAGAGAGCTTAAACGCTC 2297  
 652 AsnPheAsnArgLeuPheGluGlyArgSerSerArgLysProGluLysLeuLysThrLeu 671  
 2298 TTCTGCTACTTTAGAGAGTACAGAGAAACCCAGCTGGGTGTGGTATTCACAGATTC 2357  
 672 PheCysTyrPheArgArgValThrGluLysLysProThrGlyLeuValThrPheThrArg 691  
 2358 CAGAGCTTTGAAGATTTTCCAGAGTGGGAAAGATGTGAAAACTCTGACTCGACTGCAT 2417  
 692 GlnSerLeuGluAspPheProGluTrpGluArgCysGluLysProLeuThrArgLeuHis 711  
 2418 GTCACTTAGAAGGTACCATAGAGAAACGCCAGGCGCATGCTACAGTGTGATTTTGA 2477  
 712 ValThrTyrGluGlyThrIleGluGlyAsnGlyArgGlyMetLeuGlnValAspPheAla 731  
 2478 AACCGTTCTGGAGTGTGTAACTAGCGAGGACTTGTGCAAGAGAAATCCGCTTT 2537  
 732 AsnArgPheValGlyGlyValThrGlyAlaGlyLeuValGlnGluGluLeuArgPhe 751  
 2538 TTAATCAACCTCAGTGTGATTTTCAACGCTCTTCACTGAGTGTGATTCACATGAA 2597  
 752 LeuIleAsnProGluLeuIleValSerArgLeuPheThrGluValLeuAspHisAsnGlu 771  
 2598 TGCTTATCATCAGGTTACTGAGCAGTACAGTGAATACACAGGCTATCGCAAAACATAC 2657



Db 772 CysLeuIleIleThrGlyThrGluGlnIlyrSerGluTyrThrGlyTyrAlaGlnThrTyr 791  
QY 2658 CGCTGGCCGCGAGCCATGAAGACAGAGCGAAGAGGACGACTGGCAGAGGCGCAGACT 2717  
Db 792 ArgTrrAlaArgSerHisGluAspGlySerGluLysAspTrpGlnArgArgCysThr 811  
QY 2718 GAGATCGTCGCCATCAGCGCCCTCCACTTCAGACGCTACTCGACACGTTGTGGCCGAG 2777  
Db 812 GluIleValAlaIleAspAlaLeuHisPheArgTyrLeuAspGlnPheValProGlu 831  
QY 2778 AAGATCAGCGGAGCTTACAGGCTTACTGTGGATTCTTCGTCTGAGTTCCTCA 2837  
Db 832 LysValArgArgGluLeuAenLysAlaIlyrCysGlyPheLeuArgProGlyValProSer 851  
QY 2838 GAGAACCTGTCTGAGTGGCTACAGGAACTGGGGCTGTGGTGGCTTGGGGTGATGCT 2897  
Db 852 GluAsnLeuSerAlaValAlaThrGlyAsnTrpGlyCysGlyAlaPheGlyGlyAspAla 871  
QY 2898 AGACTAAAGCTTATACAGATCCTGGCAGCTGCTAGTCTGAGCGAGACGTGGTTAT 2957  
Db 872 ArgLeuLysAlaLeuIleGlnIleLeuAlaAlaAlaAlaGluArgAspValTyr 891  
QY 2958 TTCACTTTGGGACTCAGAACTGATCAGAGACATTTACAGATGCATATCTCTCACT 3017  
Db 892 PheThrPheGlyAspSerGluLeuMetArgAspIleTyrSerMetHisThrPheLeuThr 911  
QY 3018 GAGAGGAACTGACTGTGTGGAGAGTATATAGCTGTCTGCTACGATTTACATGAGAA 3077  
Db 912 GluArgLysLeuAspValGlyLysValTyrLysLeuLeuArgTyrTyrAsnGlu 931  
QY 3078 TGCAGAACTGCTCCACCCCGGACGACATCAAGCTTTATCCATTCATATACATGCA 3137  
Db 932 CysArgAsnCysSerThrProGlyProAspIleLysLeuTyrProPheIleTyrHisAla 951  
QY 3138 GTTCAGTCTGTACACAGACCAACCAAGCGCGGACAAAGGAGCGGG 3185  
Db 952 ValGluSerSerAlaGluThrThrAspMetProGlyGlnLysAlaGly 967  
RESULT 10  
AAU76022  
ID AAU76022 standard; protein; 968 AA.  
AC AAU76022;  
XX  
DT 08-MAY-2002 (first entry)  
XX  
DE Mouse poly (ADP-ribose) glycohydrolase (PARG) protein sequence.  
XX  
KW Mouse; poly (ADP-ribose) glycohydrolase; PARG; DNA repair; ADP-ribose;  
KW adenosine diphosphate-ribose; neoplastic disorder; adenocarcinoma;  
KW leukaemia; lymphoma; melanoma; myeloma; sarcoma; teratocarcinoma;  
KW hyperplasia; hypertrophy; reperfusion; ischaemia; heart attack; stroke;  
KW neurodegenerative disease; neurological disorder; Alzheimer's disease;  
KW Huntington's disease; Parkinson's disease.  
XX  
OS Mus musculus.  
XX  
PN US6337202-B1.  
XX  
PD 08-JAN-2002.  
XX  
PF 23-FEB-2000; 2000US-00511477.  
XX  
PR 01-MAY-1998; 98US-0083768P.  
XX  
PR 30-APR-1999; 99US-00302812.  
XX  
PA (KENT ) UNIV KENTUCKY RES FOUND.  
XX  
PI Jacobson MK, Jacobson EL, Ame J, Lin W;  
XX WPI; 2002-163240/21.  
XX  
DR N-PSDB; ABK14933.  
XX

PT Novel isolated and purified poly (ADP-ribose) glycohydrolase protein.  
PT which catalyses release of ADP-ribose from ADP ribose polymer, useful for  
XX treating neoplastic and neurological disorders, heart attack and stroke.  
PS Claim 2; Col 63-70; 81pp; English.  
XX  
CC The present invention relates to a new poly (ADP-ribose) glycohydrolase  
CC (PARG) protein which catalyses release of ADP-ribose from an ADP  
CC (adenosine diphosphate)-ribose polymer. The PARG molecule of the  
CC invention is useful for generating antibodies and can be inhibited or  
CC activated for diagnosing and treating neoplastic disorders such as  
CC adenocarcinoma, leukaemia, lymphoma, melanoma, myeloma, sarcoma,  
CC teratocarcinoma, hyperplasia and hypertrophy, reperfusion following  
CC ischaemia, heart attack, stroke, neurodegenerative diseases, neurological  
CC disorders including Alzheimer's, Huntington's and Parkinson's diseases,  
CC and related conditions. PARG levels may be enhanced to suppress DNA  
CC repair and increase the cell's susceptibility to chemotherapy drugs.  
CC Antagonists of PARG are administered to treat or prevent neoplastic  
CC disorders. The present amino acid sequence represents the mouse PARG  
CC protein of the invention. This protein is one of several PARG proteins  
CC (AAU76020-AAU76024) of the invention  
XX  
SQ Sequence 968 AA;  
Alignment Scores: 0 Length: 968  
Pred. No.: 4332.50 Matches: 820  
Score: 89.86% Conservative: 57  
Percent Similarity: 84.02% Mismatches: 90  
Best Local Similarity: 59.82% Indels: 9  
Query Match: 5 Gaps: 5  
DB:  
US-09-302-812-1 (1-4070) x AAU76022 (1-968)  
QY 258 ATGAGTCGCGCGCCGCTGTGAGCCCTGCACCAAGCGACCCGCTGGAGCGCGTGA 317  
Db 1 MetSerAlaGlyProGlyTyrGluProCysThrLys---AlaArgTyrGlyAlaAlaGly 19  
QY 318 ACTTCTCGCGCGCGCTCGGACCGCGCGAGTTCCTCGGCGAGCGAGCGCGTCTTC 377  
Db 20 ThrSerAlaProThrAlaSerAspSerArgSerPheProGlyArgGlnArgValLeu 39  
QY 378 GATTCCAGGACGCTCGGTGCGAGTTCAGGTTCGCGCTCTCTCTCAGCGTGGCGCGCTG 437  
Db 40 AspProLysAspAlaProValGlnPheArgValProProSerProAlaCysValSer 59  
QY 438 GCGCGCGCGGACAGCAGCGCGAGCGCGCCACCTCTCTTGTTCCTCAACAGAGACTATA 497  
Db 60 GlyArgAlaGlyProHisArgGlyAsnAlaThrSerPheValPheLysGlnLysThrIle 79  
QY 498 ACCAGTTGGATGGACACTAAAGGAATCAAGACGTTGAATCAGAAAGTTTGCATAGTAA 557  
Db 80 ThrThrTrpMetAspThrLysGlyProLysThrAlaGluSerGlu-----SerLys 96  
QY 558 GAAACCAACATACAGAGAGAGATCCATGATGAGTCTGTACAAAAGAGATACTTTAT 617  
Db 97 GluAsnAsnThrArgIleAspSerMetSerValGlnLysAspAsnPheTyr 116  
QY 618 CAACATACATCGAAAAATTTAGAAATGTTTCTCAGCTAGGTTTGTATGATCACCAGTT 677  
Db 117 ProHisLysValGluLysLeuGluAsnValProGlnLeuAsnLysAspLysSerProThr 136  
QY 678 GAAAAAGGTACACAGTATTGAAGCAGCATCAGACTGGGCTATCTGTAAAGTGGCAGAA 737  
Db 137 GluLysSerSerGlnTyrLeuAsnGlnGlnThrAlaSerValCysLysTrpGlnAsn 156  
QY 738 GAAGGCCACACTCAGAACGGCTTTTGGAAAGTGAACCTCCAGCGGTAACTCTGTGTACCA 797  
Db 157 GluGlyLysHisAlaGluGlnLeuLeuAlaSerGluProProAlaGlyThrProLeuPro 176  
QY 798 GAGCAGTTTCAGTAAATGCTAATCTCGATCAGTCTGCCCAAGAGATGATCAGTGCACA 857  
Db 177 LysGlnLeuSerAsnAlaAsnIleGlyGlnSerProHisThrAspAspHisSerAspThr 196



912 GluArgLysLeuAspValGlyLysValTyrLysLeuLeuArgTyrTyrAsnGluGlu 931  
 3078 TGCAGAACTGCTCCACCCCGGACACACATCAAGCTTTATCATATACCATGCA 3137  
 932 CysArgAsnCysSerThrProGlyProAspLysLeuTyrProPheIleTyrHisAla 951  
 3138 GTTCAGTCTCTGTACACAGACACACACAGCCGCGGACAAAGAGCGGG 3185  
 952 ValGluSerSerAlaGluThrThrAspMetProGlyGlnLysAlaGly 967

RESULT 11  
 AAU76013  
 ID AAU76013 standard; protein; 968 AA.  
 XX AC AAU76013;  
 XX DT 08-MAY-2002 (first entry)  
 XX DE Mouse poly (ADP-ribose) glycohydrolase (PARG) protein sequence.  
 XX KW Mouse; poly (ADP-ribose) glycohydrolase; PARG; PARG modulator; ADP-ribose;  
 KW adenosine diphosphate-ribose; DNA repair; apoptosis; neoplasia; cancer;  
 KW inherited genetic disease; myocardial infarction; vascular stroke; aging;  
 KW neurodegeneration; Huntington's disease; Parkinson's disease;  
 KW Alzheimer's disease; neurotoxicity.  
 XX OS Mus musculus.  
 XX EN US6333148-B1.  
 XX PD 25-DEC-2001.  
 XX PF 30-APR-1999; 99US-00302812.  
 XX PR 01-MAY-1998; 98US-0083768P.  
 XX PA (KENT) UNIV KENTUCKY RES FOUND.  
 XX PI Jacobson MK, Jacobson EL, Ame J, Lin W;  
 XX WPI; 2002-153820/20.  
 XX DR N-PSDB; ABK14495.  
 XX PT Screening compounds for modulation of poly (ADP-ribose) glycohydrolase,  
 XX useful potentially for treating diseases associated with DNA damage, e.g.  
 XX cancer.  
 XX PS Claim 3; Col 63-68; 80pp; English.  
 XX CC The present invention relates to a new method for screening compounds for  
 CC ability to modulate activity of an enzyme that hydrolyses ADP (adenosine  
 CC diphosphate)-ribose from an ADP-ribose polymer. The compounds are  
 CC inhibitors or activators of PARG (poly (ADP-ribose) glycohydrolase) and  
 CC are used to treat or prevent any condition associated with DNA damage  
 CC e.g. neoplasia, inherited genetic diseases, myocardial infarction,  
 CC vascular stroke, aging and neurodegeneration e.g. Huntington's,  
 CC Parkinson's or Alzheimer's diseases, or neurotoxicity generally.  
 CC Compounds identified by the new method are more effective than known  
 CC inhibitors and have fewer side effects. The present amino acid sequence  
 CC represents the mouse PARG protein of the invention. This protein is one  
 CC of several PARG proteins (AAU75799 and AAU76012-AAU76015) of the  
 XX invention  
 XX SQ Sequence 968 AA;

Alignment Scores:  
 Pred. No.: 0 Length: 968  
 Score: 4332.50 Matches: 820  
 Percent Similarity: 89.86% Conservative: 57  
 Best Local Similarity: 84.02% Mismatches: 90  
 Query Match: 59.82% Indels: 9  
 Gaps: 5

US-09-302-812-1 (1-4070) x AAU76013 (1-968)  
 QY 258 ATGAGTCCGGGCCCCGGCTGTGAGCCCTGCACCAAGCGACCCCGCTGGACGCGCTGCA 317  
 Db 1 MetSerAlaGlyProGlyTyrGluProCysThrLys---AlaArgTyrGlyAlaAlaGly 19  
 QY 318 ACTTCTCCGCGCGCTCGACCGCCGAGCTTCCCGGACGCGGACGCGGCTCTC 377  
 Db 20 ThrSerAlaProThrAlaSerAspSerArgSerPheProGlyArgGlnArgValLeu 39  
 QY 378 GATTCCAGGACCGCTCGGTGCGAGTTCAGGTCCTCCGCGCTCTCTCAGGCTGCGCCTG 437  
 Db 40 AspProLysAspAlaProValGlnPheArgValProProSerSerProAlaCysValSer 59  
 QY 438 GCGCGGCGGACAGACACCGGAGCGCCACCTCTCTTGTTCATAAAGAACTATA 497  
 Db 60 GlyArgAlaGlyProHisArgGlyAsnAlaThrSerPheValPheLysGlnLysThrIle 79  
 QY 498 ACCAGTTGGATGACACTAAAGGATCAAGACAGTTCGATCAGAAAGTTCATAGTAAA 557  
 Db 80 ThrThrTyrMetAspThrLysGlyProLysThrAlaGluSerGlu-----SerLys 96  
 QY 558 GAAACCAACAATAACAAGAGAAATCCATGATGAGTTCGTACAAAAGATACTTTAT 617  
 Db 97 GluAsnAsnAsnThrArgLysAspSerMetSerValGlnLysAspAsnPheTyr 116  
 QY 618 CAACATAACATGGAATAATAGAAAATGTTTCTCAGCTAGGTTTGTATAGTCAACAGTT 677  
 Db 117 ProHisLysValGluLysLeuGluAsnValProGlnLeuAsnLeuAspLysSerProThr 136  
 QY 678 GAAAAGGTACACAGTATTTGAAGCAGCATCAGACTCGGCTATGTGTAAGTGGCAGAT 737  
 Db 137 GluLysSerSerGlnTyrLeuAsnGlnGlnThrAlaSerValCysLysTrpGlnAsn 156  
 QY 738 GAAGGCGCACACTCAGAAACGCTTTTGGAAAGTGAACCTCCAGCGGTAACTCTGTTACCA 797  
 Db 157 GluGlyLysHisAlaGluGlnLeuLeuAlaSerGluProProAlaGlyThrProLeuPro 176  
 QY 798 GAGCAGTTTCAGTAATGCTATGTCGATCAGTCGTCGCCCAAGAGATGATCAGTGCACACA 857  
 Db 177 LysGlnLeuSerAsnAlaAsnIleGlyGlnSerProHisThrAspAspHisSerAspThr 196  
 QY 858 AATAGTGAGGAGTAGAGATAATCAGCAGTTCGATGATGATGATGATGATGATGATGATGAT 917  
 Db 197 AspHisGluGluAspArgAspAsnGlnGlnPheLeuThrProIleLysLeuAlaAsnThr 216  
 QY 918 AAGCAGACGATGGAAGATGAACAGGCGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 977  
 Db 217 LysProThrValGlyAspGlyGln-----AlaArgSerAsnCysLysCysSerGly 233  
 QY 978 GCTTGCCATCTCCAGAAAGCTGTGAGGCTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGT 1037  
 Db 234 SerArgGlnSerValLysAspCysThrGlyCysGlnGlnGluValAspValLeuPro 253  
 QY 1038 GAGGAGCCCTGTCGACACTGCTGAGGATGTGTGATGCTGAGTGAATAAATGCAAC 1097  
 Db 254 GluSerProLeuSerAspValGlyAlaGluAspIleGlyThrGlyProLysAsnAspAsn 273  
 QY 1098 AGATTGAATAGACAAAGAGTAGTCTAGGAATTCCTCCATTTGAGAAAGAAAGTGA 1157  
 Db 274 LysLeuThrGlyGlnGluSerSerLysGlyAspSerProProPheGluLysGluSerGlu 293  
 QY 1158 CCTGAGTCAACCAATGGATGTAGATAATTCAAAATAATAGTTCAGGATTCAGAGCAGAT 1217  
 Db 294 ProGluSerProMetAspValAspAsnSerSerCysGlnAspSerGluAlaAsp 313  
 QY 1218 GAAGAGACACTCCAGGTTTGTATGACACAGGAGATAGCAGTCTCTGCTCAACACAGCAAT 1277  
 Db 314 GluGluThrSerProValPheAspGluGlnAspAspArgSerSer---GlnThrAlaAsn 332  
 QY 1278 AAACCTTCAGGTTCCCAACCAAGAGAGAGTCACTGAGTTGAGAGAGCGGTCTCTGCT 1337

Db 333 LysLeuSerSerCysGlnAlaArgGluAlaAspGlyAspLeuArgLysArgTyrLeuThr 352  
 QY 1338 AAGGAGGTGAGATTCGATTCAATTTCCAAATTTGAAGAGGAGAGAGATCGAGTGAATG 1397  
 Db 353 LysGlySerGluValArgLeuHisPheGlnPheGlu--GlyGluAsnAlaGlyThr 371  
 QY 1398 AATGATGTGAATGCCAAACGACTGGAAGTACTTCTAGCTGAATGTAGAGTGCAGAAAT 1457  
 Db 372 SerAspLeuAsnAlaLysProSerGlyAsnSerSerSerLeuAsnValGluCysArgSer 391  
 QY 1458 TCTAGACATGGGAGAAAGATTTCAAAATCACAGATCAATTCATGAGAGTGCACAAA 1517  
 Db 392 SerLysGlnHisGlyLysArgAspSerLysLysLysLysLysLysLysLysLysLys 411  
 QY 1518 GCAGAGGACAAAGAAACAAATGTAAATCAAAACAAAGACAGAAAGGAGATC 1577  
 Db 412 SerGluAspArgArgLysGluGlnCysGluValArgHisGlnArgThrGluArgLysLys 431  
 QY 1578 CTTAAATACATTCACCTTCTCCAGATAGAAATGGCTGGAACTCTTATGAG 1637  
 Db 432 ProLysTyrIleProProAsnLeuProProGluLysLysLysLysLysLysLysLys 451  
 QY 1638 GAGATCAGGAGATGCCAAGGTGTGGATCGGCTGCCCTCCCTTGAGACATCTGCCAAT 1697  
 Db 452 GluMetArgLysMetProArgCysGlyLysLysLysLysLysLysLysLysLysLys 471  
 QY 1698 CACACAGTCACTATTCGGGTAGATCTTTTCGGAATAGGAGAGTTCCTAAACCTTTCCCA 1757  
 Db 472 HisThrValThrValArgValAspLeuAlaGlyGluValProLysProPhePro 491  
 QY 1758 ACATTTTAAGATTTGGGACACACAGATCTTAAGATGCTGTTTCAGACAAAC 1817  
 Db 492 ThrHisTyrLysAspLeuTrpAspAsnLysHisValLysMetProCysSerGluGlnAsn 511  
 QY 1818 TTCTACCTGTGGAAGATGAGATGTGTGAGCGAGCTGCAGGAGCGGTGGGAATCAAT 1877  
 Db 512 LeuTyrProValGluAspGluAsnGlyGluArgThrAlaGlySerArgTrpGluLeuLys 531  
 QY 1878 CAGACTGCATCTTCAACAGCTCACTCGGCCCCAGAACCTGAAGATGCTATTCGAAG 1937  
 Db 532 GlnThrAlaLeuLeuAsnLysPheThrArgProGlnAsnLeuLysAspAlaLysLeuLys 551  
 QY 1938 TACAATGTGCATTTCTAAGAAATGGGACTTTACAGCTTTGATTGATTTCTGGGATAG 1997  
 Db 552 TyrAsnValAlaTyrSerLysLysTrpAspPheThrAlaLeuValAspPheTrpAspLys 571  
 QY 1998 GTACTAGAAAGCAGAGCTCAACACTTCTATCAGTCCATCTTGTCTGATATGCTGAA 2057  
 Db 572 ValLeuGluGluAlaGluAlaGlnHisLeuTyrGlnSerLysLeuProAspMetValLys 591  
 QY 2058 ATTGCATCTGTCTGCCAAATATTTGATCCAGCCCAATACCATCTCTGAAACAGAGATG 2117  
 Db 592 IleAlaLeuCysLeuProAsnLysCysThrGlnProLysLeuLysGlnLysMet 611  
 QY 2118 AATCATTCATCAATGTACAGGACATGTCAGTCTTTAGTATGCTTCTTCTTC 2177  
 Db 612 AsnHisSerValThrMetSerGlnGluGlnIleAlaSerLeuLeuAlaAsnAlaPhePhe 631  
 QY 2178 TGCAGCTTCCAGCAGCAATGCCAAGTGAATTCAGATGATTCAGTATTCAGATATT 2237  
 Db 632 CysThrPheProArgArgAsnAlaLysMetLysSerGluTyrSerSerTyrProAspLys 651  
 QY 2238 AACTTCAATCGTTGTTGAGGACGTTTCATCAAGGAAACAGAGAGCTTAAACGCTC 2297  
 Db 652 AsnPheAsnArgLeuPheGluGlyArgSerSerArgLysProGluLysLeuLysThrLeu 671  
 QY 2298 TTCTGCTACTTTAGAGAGTACAGAGAAACCCAGTGGTGGTGGACATTCACAGA 2357  
 Db 672 PheCysTyrPheArgArgValThrGluLysLysProThrGlyLeuValThrPheThrArg 691  
 QY 2358 CAGAGTCTTGAAGATTTTCCAGAGTGGGAAAGATGTGAAACCTCCTGACTGCATGCAT 2417  
 Db 692 GlnSerLeuGluAspPheProGluTrpGluArgCysGluLysProLeuThrArgLeuHis 711

QY 2418 GTCATTACGAGGTACCATAGAGAAACCGCCAGGCGATCTACAGTGTGATTTCGA 2477  
 Db 712 ValThrTyrGluGlyThrIleGluGlyAsnGlyArgGlyMetLeuGlnValAspPheAla 731  
 QY 2478 AACCCCTTCGTGGAGGTGGTGTACCACTGAGGACTTGTGCAAGAAAGAAATCCGCTTT 2537  
 Db 732 AsnArgPheValGlyGlyValThrGlyAlaGlyLeuValGlnGluLysLeuArgPhe 751  
 QY 2538 TTAATCAACCCCTGAGTGTGATTTCACGGCTCTTTCATCTGAGTGTGATCACATGAA 2597  
 Db 752 LeuIleAsnProGluLeuIleValSerArgLeuPheThrGluValLeuAspHisAsnGlu 771  
 QY 2598 TGTCTTATCATCACAGGTACTGAGCAGTACACTGATATCACAGGCTATCCGCAAAACATAC 2657  
 Db 772 CysLeuIleIleThrGlyThrGluGlnTyrSerGluTyrThrGlyThrAlaGluThrTyr 791  
 QY 2658 CCGTGGGCGCGAGGCATGAGACAGGAGCGAAAGCGGAGCTGGCAGAGCGCAGCT 2717  
 Db 792 ArgTrpAlaArgSerHisGluAspGlySerGluLysAspAspTrpGlnArgCysThr 811  
 QY 2718 GAGATCGTGCATCGAGCCCTCCACTTCAGACGCTACCTGACGACGAGTTGTGCCGAG 2777  
 Db 812 GluIleValAlaIleAspAlaLeuHisPheArgArgTyrLeuAspGlnPheValProGlu 831  
 QY 2778 AAGATCAGAGCGGAGCTTAAACAGGCTTACTGTGATTTCTTCTGCTCGAGTTTCTTCA 2837  
 Db 832 LysValArgArgGluLeuAsnLysAlaTyrCysGlyPheLeuArgProGlyValProSer 851  
 QY 2838 GAGAACCTCTCTGAGTGTGTACAGAAACTGGGGCTGTGGTCCCTTTGGGGGTGATGCT 2897  
 Db 852 GluAsnLeuSerAlaValAlaThrGlyAsnTrpGlyCysGlyAlaPheGlyGlyAspAla 871  
 QY 2898 AGACTAAAGACCTTAAATCAGATCTCGGAGCTGTGCTGTAGCTGAGCAGACCTGGTTAT 2957  
 Db 872 ArgLeuLysAlaLeuIleGlnIleLeuAlaAlaAlaAlaGluArgAspValValTyr 891  
 QY 2958 TTCACCTTTGGGAGCTCAAACTGATGAGAGACATTTACAGATGATATCATTCCTCACT 3017  
 Db 892 PheThrPheGlyAspSerGluLeuMetArgAspIleTyrSerMetHisThrPheLeuThr 911  
 QY 3018 GAGAGAAACTGACTGTTGGAGAGTATTAAGTCTGCTAGTATTAATCAATGAGAA 3077  
 Db 912 GluArgLysLeuAspValGlyLysValTyrLysLeuLeuArgTyrTyrAsnGluGlu 931  
 QY 3078 TGCAGAAACTGCTCCACCCCGGACCCAGACATCAAGCTTTATCCATTCATATACCATGCA 3137  
 Db 932 CysArgAsnCysSerThrProGlyProAspIleLysLeuTyrProPheIleTyrHisAla 951  
 QY 3138 GTTGAGTCTCTGTACACAGCACCAACCGCGGAGCAAGAGACAGCGGG 3185  
 Db 952 ValGluSerSerAlaGluThrThrAspMetProGlyGlnLysAlaGly 967  
 RESULT 12  
 ABG72280  
 ID ABG72280 standard; protein; 968 AA.  
 XX  
 AC ABG72280;  
 XX  
 XX 13-MAR-2003 (first entry)  
 XX  
 DE Murine poly (ADP-ribose) glycohydrolase (PARG) enzyme.  
 XX Murine; poly (ADP-ribose) glycohydrolase; PARG; PARG expression;  
 KW cellular response; DNA damage; neoplastic disorder inducing agent;  
 KW oxidative stress; neoplastic disorder; myocardial infarction;  
 KW Parkinson's disease; neurodegenerative disorder; Alzheimer's disease;  
 KW vascular stroke; Huntington's disease; inborn genetic error;  
 KW reperfusion; ischaemia; aging; neurotoxicity; PARG activity; mouse;  
 KW cytosolic; neuroprotective; neurotropic; antiparkinsonian; cardiant;  
 KW vasotrophic; anticonvulsant; cerebroprotective; enzyme.  
 XX  
 OS Mus musculus.

XX US2002132328-A1.  
 PN 19-SEP-2002.  
 PD 09-OCT-2001; 2001US-00973451.  
 PP 01-MAY-1998; 98US-0083768P.  
 PR 30-APR-1999; 99US-00302812.  
 XX (JACO) JACOBSON M K.  
 PA (JACO) JACOBSON E L.  
 PA (AMEJ) AME J.  
 PA (LINW) LIN W.  
 XX Jacobson MK, Jacobson EL, Ame J, Lin W;  
 PI WPI; 2003-155895/15.  
 DR N-PSDB; ABX14479.  
 DR  
 XX New nucleic acid molecule encoding a polypeptide with poly(ADP-ribose)  
 PT glycohydrolase activity, for preventing, treating, or ameliorating a  
 PT disease condition, e.g. neoplastic disorder, myocardial infarction or  
 PT vascular stroke.  
 XX  
 PS Claim 28; Fig 16; 86pp; English.  
 XX The present invention relates to the isolation of poly(ADP-ribose)  
 CC glycohydrolase (PARG) from several species, and the polynucleotide  
 CC sequences encoding them. Methods for inhibiting PARG expression or  
 CC overexpressing PARG are also disclosed. PARG is involved in the cellular  
 CC response to DNA damage, and is associated with the body's response to  
 CC neoplastic disorder inducing agents and oxidative stress. The  
 CC polynucleotide sequences encoding PARG and PARG modulators are useful for  
 CC preventing, treating, or ameliorating diseases such as neoplastic  
 CC disorders, myocardial infarction, vascular stroke, neurodegenerative  
 CC disorders (e.g. Alzheimer's disease, Parkinson's disease or Huntington's  
 CC disease), inborn genetic errors, reperfusion following ischaemia, aging,  
 CC and neurotoxicity. The polynucleotide sequences are also useful in gene  
 CC therapy. The methods are useful for identifying an agent that modulates  
 CC PARG activity, identifying a mutant PARG allele in an individual, or  
 CC screening candidate molecules for PARG modulating activity. The present  
 CC sequence represents murine PARG enzyme  
 XX Sequence 968 AA;  
 SQ  
 Alignment Scores:  
 Pred. No.: 0 Length: 968  
 Score: 4332.50 Matches: 820  
 Percent Similarity: 89.86% Conservative: 57  
 Best Local Similarity: 84.02% Mismatches: 90  
 Query Match: 59.82% Indels: 9  
 DB: 6 Gaps: 5  
 US-09-302-812-1 (1-4070) x ABG72280 (1-968)  
 QY 258 ATGATGTCGGGCGCGGCTGTGAGCCCTGCACCAAGCGACCCCGCTGGACCGCGCTGCA 317  
 DB 1 MetSerAlaGlyProGlyTrpGluProCysThrLys--AlaArgTrpGlyAlaGly 19  
 QY 318 ACTTCTCCGCGCCCGCTCGGACGCGCGAGCTTCCCGCGGACGAGCGCGGCTC 377  
 DB 20 ThrSerAlaProThrAlaSerArgSerPheProGlyArgGlnArgValLeu 39  
 QY 378 GATTCCAGACGCTCGGTGAGTTGAGGTCGGGTCGGGTCGGGTCGGGTCGGGTCGGGTCG 437  
 DB 40 AspProLysAspAlaProValGlnPheArgValProProSerSerProAlaCysValSer 59  
 QY 438 GCGCGGCGGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 497  
 DB 60 GlyArgAlaGlyProHisArgGlyAsnAlaThrSerPheValPheLysGlnLysThrIle 79  
 QY 498 ACCAGTTGGATGGACACTAAAGGAATCAAGACAGATTGAATCAAGAAAGTTGGCATGATA 557

DB 80 ThrThrTrpMetAspThrLysGlyProLysThrAlaGluSerGlu-----SerLys 96  
 QY 558 GAAACCAACAATCAAGAGAGAAATCCATGATGAGTTCGTACAAAAGATACTTTTAT 617  
 DB 97 GluAsnAsnAsnThrArgIleAspSerMetMetSerSerValGlnLysAspAsnPheTyr 116  
 QY 618 CAACATAACATGAAATTTAGAAATGTTCTCAGCTAGGTTTGTATAGTCACCAAGTT 677  
 DB 117 ProHisLysValGluLysLeuGluAsnValProGlnLeuAsnLeuAspLysSerProThr 136  
 QY 678 GAAAAGGTACACAGATTTTGAAGCAGCATCAGACTGCGCTATGTGTAAGTGGCAGAA 737  
 DB 137 GluLysSerSerGlnLysLeuAsnGlnGlnThrAlaSerValCysLysTyrGlnAsn 156  
 QY 738 GAAGGGCCACACTCAGAACCGCTTTTGAAGTGAACCTCAGCGGTAACTCTGGTACCA 797  
 DB 157 GluGlyLysHisAlaGluGlnLeuAlaSerGluProAlaGlyThrProLeuPro 176  
 QY 798 GAGCAGTTCAGTATGCTATGCTCATGCTCAGTCGTCCTCCCAAGGATGATCAGTGCACA 857  
 DB 177 LysGlnLeuSerAsnAlaAsnIleGlyGlnSerProHisThrAspAspHisSerAspThr 196  
 QY 858 AATAGTCAGGAGTACAGATAATCAGCAGTCTTTTGCACATGTGATAAGCTTGCAGATGCA 917  
 DB 197 AspHisGluGluAspArgAspAsnGlnGlnPheLeuThrProLysLeuAlaAsnThr 216  
 QY 918 AAGCAGACGATGGAAGATGAACAGGCGAGAGAACCCAGACCCAGAGTGTGGCAG 977  
 DB 217 LysProThrValGlyAspGlyGln-----AlaArgSerAsnCysLysCysSerGly 233  
 QY 978 GCTTGGCCATCTCGCAAGACCTGTGAGGCTGTGAGGAGTGTGAGGAGGAGGAGGAGTGTG 1037  
 DB 234 SerArgGlnSerValLysAspCysThrGlyCysGlnGlnGluValAspValLeuPro 253  
 QY 1038 GAGAGCCCTTGTGCGACACTGCTGTGAGATGCTGTGATGCTGTGAGTGTGAGTGTGAGTGTG 1097  
 DB 254 GluSerProLysSerAspValGlyAlaGluAspIleGlyThrGlyProLysAsnAspAsn 273  
 QY 1098 AGATTGATACACAAAGATGCTAGGAAATCTCTCCATTTGAGAAAGAAAGTGA 1157  
 DB 274 LysLeuThrGlyGlnGluSerSerLeuGlyAspSerProPheGluLysGluSerGlu 293  
 QY 1158 CTTGAGTCCCAATGAGTGTAGATAATTCCTCAAAATAGTGTGAGGATTCAGGATTCAGAAC 1217  
 DB 294 ProGluSerProMetAspValAspAsnSerArgAsnSerCysGlnAspSerGluAlaAsp 313  
 QY 1218 GAAGACACAAGTCCAGGTTTGTGATGAACAGGAGATAGCTTCTGCTCAACACAGCAAT 1277  
 DB 314 GluGluThrSerProValPheAspGluGlnAspAspArgSerSer---GlnThrAlaAsn 332  
 QY 1278 AAACCTTCAAGTTCCCAACAGAGAGAGTGCACACTGAGTGTGAGGAGGAGGAGTGTGCT 1337  
 DB 333 LysLeuSerSerCysGlnAlaArgGluAlaAspGlyLeuArgLysArgTyrLeuThr 352  
 QY 1338 AAGGAGGTGAGTTCGATTCATTTCAATTTGAAGGAGGAGGAGGAGTGTGAGTGTGAGTGTG 1397  
 DB 353 LysGlySerGluValArgLeuHisPheGlnPheGlu---GlyGluAsnAlaGlyThr 371  
 QY 1398 AATGATGTGAATGCCAAACACGACTCGAAGTACTTCTAGCCTGAATGTAGTGTGAGTGTG 1457  
 DB 372 SerAspLeuAsnAlaLysProSerGlyAsnSerSerLeuAsnValGluCysArgSer 391  
 QY 1458 TCTAAGCAACATGGAGGAAAGGATTTCTAAATCAAGATCATTTCTCATGAGTGTGCCAAA 1517  
 DB 392 SerLysGlnHisGlyLysArgAspSerLysIleThrAspHisPheMetArgIleSerLys 411  
 QY 1518 GCAGAGGCAAAAGAAAGAAACAATGTGAATGAACATCAAGAAAGCAAGAAAGGAGATC 1577  
 DB 412 SerGluAspArgArgLysGluGlnCysGluValArgHisGlnArgThrGluArgLysIle 431  
 QY 1578 CCTAATAATCTCCACCTCCTCTCTCCAGATAAGAAATGCTGTGGAACTCCTATTGAG 1637

Db 432 ProLysTyrIleProProAsnLeuProProGluLysLysTriLeuGlyThrProIleGlu 451  
QY 1638 GAGATGAGGAGATGCCAAGGTGTGGATGCCGGCTGCTCCCTTGAGACCATCTGCCAAT 1697  
Db 452 GluMetArgLysMetProArgCysGlyIleHisLeuProSerLeuArgProSerAlaSer 471  
QY 1698 CACACAGTCACTATTCCGGTGTAGATCTTTTCCGATAGGAGAGAGTCTCTAAACCTTCCCA 1757  
Db 472 HisThrValThrValArgValAspLeuLeuArgAlaGlyGluValProLysProPhePro 491  
QY 1758 ACACATTTTAAAGATTGTGGGACACAAAGCATGTTAAGATGCTTTGTTTCAGAACAAAAC 1817  
Db 492 ThrHisTyrLysAspLeuTriAspAsnLysHisValLysMetProCysSerGluGlnAsn 511  
QY 1818 TTGTACCTGTGGAAGATGAGATGTGCGAGCTGCGAGCGCGGTGGGACTCAT 1877  
Db 512 LeuTyrProValGluAspGluAsnGlyGluArgThrAlaGlySerArgTrpGluLeuIle 531  
QY 1878 CAGACTGCACTTCTCAACAGGCTCACTCCGCCCCAGCACTGAAGGATGCTATCTGAAG 1937  
Db 532 GlnThrAlaLeuLeuAsnLysPheThrArgProGlnAsnLeuLysAspAlaIleLeuLys 551  
QY 1938 TACAATGTGCAATTTCTAAGAAATGGACTTTTACAGCTTTTGAATTTCTGGGATAAG 1997  
Db 552 TyrAsnValAlaTyrSerLysLysTrpAspPheThrAlaLeuValAspPheTrpAspLys 571  
QY 1998 GTACTAGAGAAAGCAGAGCTCAACACTGTGTATCAGTCCATCTTGTGCTGATATGGTGA 2057  
Db 572 ValLeuGluGluAlaGluAlaGlnHisLeuTyrGlnSerIleLeuProAspMetValLys 591  
QY 2058 ATTGCACTGTCTGCGCAATATTTGTACCACGCCAATACACTCTCTGAAACAGAAAGATG 2117  
Db 592 IleAlaLeuCysLeuProAsnIleCysThrGlnProIleProLeuLeuLysGlnLysMet 611  
QY 2118 AATCATTTCCATCAATGTCCAGGAAACAGATGGCCAGTCTTTTAGCTAATGCTTCTTC 2177  
Db 612 AsnHisSerValThrMetSerGlnGluGlnIleAlaSerLeuLeuAlaAsnAlaPhePhe 631  
QY 2178 TGCACTTCCAGCAGCAATGTCAGATCAATCAATCAGATGATTCAGATTTCCAGATATT 2237  
Db 632 CysThrPheProArgAsnAlaLysMetLysSerGluTyrSerSerTyrProAspIle 651  
QY 2238 AACTTCAATCGGTGTTTGAAGCAGCTTCATCAAGAAACAGAGAACTTAAACAGCTC 2297  
Db 652 AsnPheAsnArgLeuPheGluArgSerSerArgLysProGluLysLeuLysThrLeu 671  
QY 2298 TTCTGCTTACTTTAGAAGAGTACAGAGAAACCCCACTGGTGTGTGACATTCACAGA 2357  
Db 672 PheCysTyrPheArgArgValThrGluLysLysProThrGlyLeuValThrPheThrArg 691  
QY 2358 CAGAGCTTGAAGATTTCCAGAGTGGGAAGATGTGAAAACCTCTGACTCGACTGCAT 2417  
Db 692 GlnSerLeuGluAspPheProGluTrpGluArgCysGluLysProLeuThrArgLeuHis 711  
QY 2418 GTCACTTACCAAGTACCATAGAGAAACCGCCAGGGCATGTCTACAGTGTGATTTTGCA 2477  
Db 712 ValThrTyrGluGlyThrIleGluGlyAsnGlyArgGlyMetLeuGlnValAspPheAla 731  
QY 2478 AACCGTTTCTGTTGAGGTGTGTAAACAGTGCAGGACTTGTGCAAGAAACATCCGCTTT 2537  
Db 732 AsnArgPheValGlyGlyGlyValThrGlyAlaGlyLeuValGlnGluLeuIleArgPhe 751  
QY 2538 TTAATCAACCTTGATGTTGTTTTCAGGCTCTTCACTGAGTGTGCTGATCACAATGAA 2597  
Db 752 LeuIleAsnProGluLeuIleValSerArgLeuPheThrGluValLeuAspHisAsnGlu 771  
QY 2598 TGTCTTATCATCACAGTACTGACGAGTACAGTCAATACAGGCTATGCGAAACATAC 2657  
Db 772 CysLeuIleIleThrGlyThrGluGlnTyrSerGluTyrThrGlyTyrAlaGluThrTyr 791  
QY 2658 CGCTGGGCGCCGAGCCATGAAAGACAGGCGAAAGGACGACTGCGAGAGCGCACCAT 2717  
Db 792 ArgTrpAlaArgSerHisGluAspGlySerGluLysAspAspTrpGlnArgCysThr 811

QY 2718 GAGATCGTCGCATCGACGCCCTCCACTTCAGACGCTACCTCGACGAGTTTGTGCCGAG 2777  
Db 812 GluIleValAlaIleAspAlaLeuHisPheArgArgTyrLeuAspGlnPheValProGlu 831  
QY 2778 AAGATCAGACGGGAGCTTAAAGCTTACTGTGTGATTTCTTCGCTCGCTGCTTCTTCA 2837  
Db 832 LysValArgArgGluLeuAsnLysAlaTyrCysGlyPheLeuArgProGlyValProSer 851  
QY 2838 GAGACCTGTCTGCGAGTGGCTACAGMAACTCGGGCTGTGGTTCGGGGTGATGCT 2897  
Db 852 GluAsnLeuSerAlaValAlaThrGlyAsnTrpGlyCysGlyAlaPheGlyGlyAspAla 871  
QY 2898 AGACTAAAGACCTTAATACAGATCTCGCAGCTGTGTAGCTGAGCGAGACGCTGTTAT 2957  
Db 872 ArgLeuLysAlaLeuIleGlnIleLeuAlaAlaAlaGluArgAspValValTyr 891  
QY 2958 TTCACTTTTGGGACTTCAGACTGATGAGAGACATTTACAGATGATCATTCCTCACT 3017  
Db 892 PheThrPheGlyAspSerGluLeuMetArgAspIleTyrSerMetHisThrPheLeuThr 911  
QY 3018 GAGAGAAACTGACTGTTGGAGAGTATATAAGCTGCTGCTACGATATACATGAAGAA 3077  
Db 912 GluArgLysLeuAspValGlyLysValTyrLysLeuLeuArgTyrTyrAsnGluGlu 931  
QY 3078 TGCAGAACTGTCTCCACCCCGGACCCAGACATCAAGCTTTATCCATTCATACCATGCA 3137  
Db 932 CysArgAsnCysSerThrProGlyProAspIleLysLeuTyrProPheIleTyrHisAla 951  
QY 3138 GTTGAAGTCTGTACAGACACCCAAACAGCCGGGCAAGAGCGGG 3185  
Db 952 ValGluSerSerAlaGluThrThrAspMetProGlyGlnLysAlaGly 967

## RESULT 13

ABBS59491 standard; protein; 768 AA.

XX AC ABBS59491;

XX DT 26-MAR-2002 (first entry)

XX DE Drosophila melanogaster polypeptide SEQ ID NO 5265.

XX KW Drosophila; developmental biology; cell signalling; insecticide; pharmaceutical.

XX OS Drosophila melanogaster.

XX FN WO200171042-A2.

XX PD 27-SEP-2001.

XX PF 23-MAR-2001; 2001WO-US009231.

XX PR 23-MAR-2000; 2000US-0191637P.

XX PR 11-JUL-2000; 2000US-00614150.

XX PA (PEKE ) PE CORP NY.

XX XX Venter JC, Adams M, Li PWD, Myers EW;

XX XX WPI; 2001-656860/75.

XX DR N-PSDB; ABL03594.

XX DR New isolated nucleic acid detection reagent for detecting 1000 or more genes from Drosophila and for elucidating cell signaling and cell-cell interactions.

XX PS Disclosure; SEQ ID NO 5265; 21pp + Sequence Listing; English.

XX CC The invention relates to an isolated nucleic acid detection reagent capable of detecting 1000 or more genes from Drosophila. The invention is useful in developmental biology and in elucidating cell signalling and











582 LysLysLysLeuTyrAspPheIle-----LysGluGluLeuLysLysValArgAsp 598

3165 CAGCCGGACAAAGGACGGGGCC 3188

599 ValProGlyGluGlyAlaSerAla 606

Search completed: May 26, 2004, 16:45:16  
Job time : 297.648 secs

Db 231 ValProLeuLeuLysHisHisLysAsnAlaSerLeuSerLeuGlnGlnIleSer 250  
QY 2154 AGTCTTTTAGCTAATGCTTTCTTCGACGCTTTCCACGACCAATGCC---AGATGAAA 2210  
Db 251 CysLeuLeuAlaAsnAlaPheLeuCysThrPheProArgArgAsnThrLeuLysArgLys 270  
QY 2211 TCAGAGTATCCAGTATCCAGATATTAATCTCAATCGGTGTTTGAAGGACGTCATCA 2270  
Db 271 SerGluTyrSerThrPheProAspIleAsnPheAsnArgLeuTyrGlnSerThrGlyPro 290  
QY 2271 AGGAAACCCAGAGAGCTTAAACGCTCTCTGCTACTTTAGAACAGTC-----ACAGAG 2324  
Db 291 AlaValLeuGluLysLeuLysCysIleMethIstYrPheArgArgValCysProThrGlu 310  
QY 2325 AAAAAA-----CCCACTGGTGTGTGACATTCACAAACAGAGT-----CTT 2366  
Db 311 ArgAspAlaSerAsnValProThrGlyValThrPheValArgArgSerGlyLeuPro 330  
QY 2367 GAAGATTTCCAGAGTGGGAAGATGTGAATACTCCTG-----ACTCGACTGATGTC 2420  
Db 331 GluHisLeuIleAspTrpSerGlnSerAlaAlaProLeuGlyAspValProLeuHisVal 350  
QY 2421 ACTTACGAGGTACCATAGAGAAACGCCAGGCGATGCTACAGTGGATTTTCAAAAC 2480  
Db 351 AspAlaGluGlyThrIleGluAspGluGlyIleGlyLeuLeuGlnValAspPheAlaAsn 370  
QY 2481 CGTTTCGTTGGAGGTGGTGAACACAGTGCAGGACTTGTCAAGAAAGAAATCCGCTTTTA 2540  
Db 371 LysTyrLeuGlyGlyGlyValLeuGlyHisGlyCysValGlnGluIleArgPheVal 390  
QY 2541 ATCAACCTCGAGTTCATGTTTCACGGCTCTTCACAGCTGCTGATCAATGAATGT 2600  
Db 391 IleCysProGluLeuValGlyLysLeuPheThrGluCysLeuArgProPheGluAla 410  
QY 2601 CTTATCATCAGGTACTCAGCAGTACATGAATACACAGGCTATGCCGAAACATACCGC 2660  
Db 411 LeuValMetLeuGlyAlaGluArgTyrSerAsnTyrThrGlyTyrAlaGlySerPheGlu 430  
QY 2661 TGGGCCGCGAGCATCAGACAGGAGGAAAGGACGACTGGCAGAGCGGCACGACTGAG 2720  
Db 431 TrpSerGlyAsnPheGluAspSerThrProArgAspSerSerGlyArgArgGlnThrAla 450  
QY 2721 ATGTCGCCATCAGCCCTCCATTCACAGCGCTACCTCGACAGTTTGTGCCGAGAGAAG 2780  
Db 451 IleValAlaIleAspAlaLeuHisPheAlaGlnSerHisGlnTyrArgGluAspLeu 470  
QY 2781 ATCAGACGGGAGCTTAAACAGGCTTACTGTGGATTCTTCGT-----2822  
Db 471 MetGluArgGluLeuAsnLysAlaTyrIleGlyPheValHisTrpMetValThrProPro 490  
QY 2823 CTGGAGATTCTTCAGAGAACCTGTCTGCGAGTGGCTACAGGAACTGGGGCTGTGGTGC 2882  
Db 491 ProGly-----ValAlaThrGlyAsnTrpGlyCysGlyAla 502  
QY 2883 TTTGGGGGTGATCTAGACTAAAGCTTAATACAGATCCTGCGAGCTGCTGAGCTGAG 2942  
Db 503 PheGlyAspSerTyrLeuLysAlaLeuLeuGlnLeuMetValCysAlaGlnLeuGly 522  
QY 2943 CGAGAGCTGTGTTATTTACCTTTGGGACTCAGAACTGATGAGAGACATTTACAGCATG 3002  
Db 523 ArgProLeuAlaTyrTyrThrPheGlyAsnValGluPheArgAspPheHisGluMet 542  
QY 3003 CATACATTCCTCAGTACAGAGAACTGCTGTTGGAGAGATATATAGCTGCTGCTACGA 3062  
Db 543 TrpLeuLeuPheArgAsnAspGlyThrThrValGlnGlnLeuTrpSer---IleLeuArg 561  
QY 3063 TATTACAAATGAGAAATGACAGAACTGCTCCACCCCGGACCA-----3104  
Db 562 SerTyrSerArgLeuIleLysGluLysSerSerLysGluProArgGluAsnLysAlaSer 581  
QY 3105 GACATCAGCTTTATCCATTATCATATACCATGCTGAGTTGCTGCTGTACACAGACCACCAAC 3164

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 26, 2004, 18:38:05 ; Search time 0.352479 Seconds  
(without alignments)  
1611.120 Million cell updates/sec

Title: US-09-302-812-13  
Perfect score: 53  
Sequence: 1 FLINPELIVSR 11

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents AA:\*  
1: /cgn2\_6/prodata/2/iaa/5A\_COMB.pep:\*  
2: /cgn2\_6/prodata/2/iaa/5B\_COMB.pep:\*  
3: /cgn2\_6/prodata/2/iaa/6A\_COMB.pep:\*  
4: /cgn2\_6/prodata/2/iaa/6B\_COMB.pep:\*  
5: /cgn2\_6/prodata/2/iaa/PCUTUS\_COMB.pep:\*  
6: /cgn2\_6/prodata/2/iaa/backfile1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	53	100.0	11	4 US-09-302-812-13	Sequence 13, Appl
2	53	100.0	11	4 US-09-511-477-13	Sequence 13, Appl
3	53	100.0	11	4 US-09-511-507-13	Sequence 13, Appl
4	53	100.0	968	4 US-09-302-812-6	Sequence 6, Appl
5	53	100.0	968	4 US-09-511-477-6	Sequence 6, Appl
6	53	100.0	968	4 US-09-511-507-6	Sequence 6, Appl
7	53	100.0	977	4 US-09-302-812-2	Sequence 2, Appl
8	53	100.0	977	4 US-09-511-477-2	Sequence 2, Appl
9	53	100.0	977	4 US-09-511-507-2	Sequence 2, Appl
10	52	98.1	976	4 US-09-302-812-4	Sequence 4, Appl
11	52	98.1	976	4 US-09-511-477-4	Sequence 4, Appl
12	52	98.1	976	4 US-09-511-507-4	Sequence 4, Appl
13	41	77.4	252	4 US-09-198-452A-694	Sequence 694, App
14	37	69.8	1336	1 US-08-026-138B-4	Sequence 4, Appl
15	37	69.8	1336	2 US-08-231-193A-58	Sequence 58, Appl
16	37	69.8	1336	2 US-08-486-273A-58	Sequence 58, Appl
17	37	69.8	1336	3 US-08-940-086A-58	Sequence 58, Appl
18	37	69.8	1336	4 US-08-940-035A-58	Sequence 58, Appl
19	37	69.8	1336	4 US-08-935-105A-58	Sequence 58, Appl
20	37	69.8	1336	4 US-09-648-797-58	Sequence 58, Appl
21	37	69.8	1336	4 US-09-386-123-58	Sequence 58, Appl
22	36	67.9	123	4 US-09-328-352-5925	Sequence 5925, App
23	35	66.0	358	4 US-08-858-207A-398	Sequence 398, App
24	34	64.2	134	4 US-09-489-039A-11592	Sequence 11592, A
25	34	64.2	409	4 US-09-540-236-2952	Sequence 2952, App
26	34	64.2	1065	3 US-09-412-545-2	Sequence 2, Appl
27	33	62.3	88	4 US-09-328-352-7192	Sequence 7192, App

28	33	62.3	176	4	US-09-107-532A-6622	Sequence 6622, Ap
29	33	62.3	313	4	US-09-551-826D-14	Sequence 14, Appl
30	33	62.3	400	4	US-09-543-681A-4587	Sequence 4587, Ap
31	33	62.3	452	4	US-09-198-452A-853	Sequence 853, App
32	33	62.3	883	4	US-09-489-039A-11249	Sequence 11249, A
33	32	60.4	169	2	US-08-895-939-4	Sequence 4, Appli
34	32	60.4	169	3	US-09-188-820-4	Sequence 4, Appli
35	32	60.4	172	4	US-09-543-681A-7238	Sequence 7238, Ap
36	32	60.4	176	4	US-09-543-681A-6922	Sequence 6922, Ap
37	32	60.4	247	4	US-09-134-000C-6345	Sequence 6345, Ap
38	32	60.4	264	1	US-08-188-582-26	Sequence 26, Appl
39	32	60.4	264	1	US-08-646-715-26	Sequence 26, Appl
40	32	60.4	325	2	US-09-018-576-3	Sequence 3, Appli
41	32	60.4	325	2	US-09-018-576-12	Sequence 12, Appl
42	32	60.4	325	3	US-09-248-137-3	Sequence 3, Appli
43	32	60.4	325	3	US-09-248-137-12	Sequence 12, Appl
44	32	60.4	422	4	US-09-252-991A-27513	Sequence 27513, A
45	32	60.4	501	4	US-09-687-360-2	Sequence 2, Appli

## ALIGNMENTS

RESULT 1  
US-09-302-812-13  
; Sequence 13, Application US/09302812B  
; Patent No. 6333148  
; GENERAL INFORMATION:  
; APPLICANT: JACOBSON, Myron K.  
; APPLICANT: JACOBSON, Elaine L.  
; APPLICANT: AM, Jean-Christophe  
; APPLICANT: LIN, Winston  
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE (PARG) ENZ  
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIVE  
; TITLE OF INVENTION: THEREWITH  
; FILE REFERENCE: NIAD 201  
; CURRENT APPLICATION NUMBER: US/09/302,812B  
; CURRENT FILING DATE: 1999-04-30  
; EARLIER APPLICATION NUMBER: 60/083,768  
; EARLIER FILING DATE: 1998-05-01  
; NUMBER OF SEQ ID NOS: 38  
; SEQ ID NO 13  
; LENGTH: 11  
; TYPE: PRT  
; ORGANISM: Bos taurus  
; FEATURE:  
US-09-302-812-13

Query Match 100.0%; Score 53; DB 4; Length 11;  
Best Local Similarity 100.0%; Pred. No. 0.00016;  
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FLINPELIVSR 11  
|||||  
Db 1 FLINPELIVSR 11

RESULT 2  
US-09-511-477-13  
; Sequence 13, Application US/09511477  
; Patent No. 6337202  
; GENERAL INFORMATION:  
; APPLICANT: JACOBSON, Myron K.  
; APPLICANT: JACOBSON, Elaine L.  
; APPLICANT: AME, Jean-Christophe  
; APPLICANT: LIN, Winston  
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE (PARG) ENZ  
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIVE  
; TITLE OF INVENTION: THEREWITH  
; FILE REFERENCE: NIAD 201  
; CURRENT APPLICATION NUMBER: US/09/511,477  
; CURRENT FILING DATE: 2000-02-23  
; PRIOR APPLICATION NUMBER: 09/302,812

; PRIOR FILING DATE: 1999-04-30  
; NUMBER OF SEQ ID NOS: 38

; SEQ ID NO 13

; LENGTH: 11

; TYPE: PRT

; ORGANISM: Bos taurus

; FEATURE:

US-09-511-477-13

Query Match 100.0%; Score 53; DB 4; Length 11;  
Best Local Similarity 100.0%; Pred. No. 0.00016;  
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FLINPELIVSR 11

Db 1 FLINPELIVSR 11

#### RESULT 3

US-09-511-507-13

; Sequence 13, Application US/09511507

; Patent No. 6395543

; GENERAL INFORMATION:

; APPLICANT: JACOBSON, Myron K.

; APPLICANT: JACOBSON, Elaine L.

; APPLICANT: AME, Jean-Christophe

; APPLICANT: LIN, Winston

; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE (PARG) EN

; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV

; TITLE OF INVENTION: THEREWITH

; FILE REFERENCE: NIAD 201

; CURRENT APPLICATION NUMBER: US/09/511,507

; CURRENT FILING DATE: 2000-02-23

; PRIOR APPLICATION NUMBER: 09/302,812

; PRIOR FILING DATE: 1999-04-30

; NUMBER OF SEQ ID NOS: 38

; SEQ ID NO 13

; LENGTH: 11

; TYPE: PRT

; ORGANISM: Bos taurus

; FEATURE:

US-09-511-507-13

Query Match 100.0%; Score 53; DB 4; Length 11;  
Best Local Similarity 100.0%; Pred. No. 0.00016;  
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FLINPELIVSR 11

Db 1 FLINPELIVSR 11

#### RESULT 4

US-09-302-812-6

; Sequence 6, Application US/09302812B

; Patent No. 6333148

; GENERAL INFORMATION:

; APPLICANT: JACOBSON, Myron K.

; APPLICANT: JACOBSON, Elaine L.

; APPLICANT: AM, Jean-Christophe

; APPLICANT: LIN, Winston

; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE (PARG) EN

; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV

; TITLE OF INVENTION: THEREWITH

; FILE REFERENCE: NIAD 201

; CURRENT APPLICATION NUMBER: US/09/302,812B

; CURRENT FILING DATE: 1999-04-30

; EARLIER APPLICATION NUMBER: 60/083,768

; EARLIER FILING DATE: 1998-05-01

; NUMBER OF SEQ ID NOS: 38

; SEQ ID NO 6

; LENGTH: 968

; TYPE: PRT

; ORGANISM: Mus musculus

; FEATURE:

US-09-302-812-6

Query Match 100.0%; Score 53; DB 4; Length 968;  
Best Local Similarity 100.0%; Pred. No. 0.036;  
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FLINPELIVSR 11

Db 751 FLINPELIVSR 761

#### RESULT 5

US-09-511-477-6

; Sequence 6, Application US/09511477

; Patent No. 6337202

; GENERAL INFORMATION:

; APPLICANT: JACOBSON, Myron K.

; APPLICANT: JACOBSON, Elaine L.

; APPLICANT: AME, Jean-Christophe

; APPLICANT: LIN, Winston

; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE (PARG) EN

; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV

; TITLE OF INVENTION: THEREWITH

; FILE REFERENCE: NIAD 201

; CURRENT APPLICATION NUMBER: US/09/511,477

; CURRENT FILING DATE: 2000-02-23

; PRIOR APPLICATION NUMBER: 09/302,812

; PRIOR FILING DATE: 1999-04-30

; NUMBER OF SEQ ID NOS: 38

; SEQ ID NO 6

; LENGTH: 968

; TYPE: PRT

; ORGANISM: Mus musculus

; FEATURE:

US-09-511-477-6

Query Match 100.0%; Score 53; DB 4; Length 968;  
Best Local Similarity 100.0%; Pred. No. 0.036;  
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FLINPELIVSR 11

Db 751 FLINPELIVSR 761

#### RESULT 6

US-09-511-507-6

; Sequence 6, Application US/09511507

; Patent No. 6395543

; GENERAL INFORMATION:

; APPLICANT: JACOBSON, Myron K.

; APPLICANT: JACOBSON, Elaine L.

; APPLICANT: AME, Jean-Christophe

; APPLICANT: LIN, Winston

; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE (PARG) EN

; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV

; TITLE OF INVENTION: THEREWITH

; FILE REFERENCE: NIAD 201

; CURRENT APPLICATION NUMBER: US/09/511,507

; CURRENT FILING DATE: 2000-02-23

; PRIOR APPLICATION NUMBER: 09/302,812

; PRIOR FILING DATE: 1999-04-30

; NUMBER OF SEQ ID NOS: 38

; SEQ ID NO 6

; LENGTH: 968

; TYPE: PRT

; ORGANISM: Mus musculus

; FEATURE:

US-09-511-507-6

Query Match 100.0%; Score 53; DB 4; Length 968;

Best Local Similarity 100.0%; Pred. No. 0.036; Mismatches 0; Indels 0; Gaps 0;  
Matches 11; Conservative 0;

QY 1 FLINPELIVSR 11  
DB 751 FLINPELIVSR 761

RESULT 7  
US-09-812-2  
; Sequence 2, Application US/09302812B  
; Patent No. 633148  
; GENERAL INFORMATION:  
; APPLICANT: JACOBSON, Myron K.  
; APPLICANT: JACOBSON, Elaine L.  
; APPLICANT: AM, Jean-Christophe  
; APPLICANT: LIN, Winston  
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY(ADP-RIBOSE) GLYCOHYDROLASE (PARG) ENZ  
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIVE  
; TITLE OF INVENTION: THEREWITH  
; FILE REFERENCE: NIAD 201  
; CURRENT APPLICATION NUMBER: US/09/302,812B  
; CURRENT FILING DATE: 1999-04-30  
; EARLIER APPLICATION NUMBER: 60/083,768  
; EARLIER FILING DATE: 1998-05-01  
; NUMBER OF SEQ ID NOS: 38  
; SEQ ID NO 2  
; LENGTH: 977  
; TYPE: PRT  
; ORGANISM: Bos taurus  
; FEATURE:  
US-09-302-812-2

Query Match 100.0%; Score 53; DB 4; Length 977;  
Best Local Similarity 100.0%; Pred. No. 0.037; Mismatches 0; Indels 0; Gaps 0;  
Matches 11; Conservative 0;

QY 1 FLINPELIVSR 11  
DB 760 FLINPELIVSR 770

RESULT 8  
US-09-511-477-2  
; Sequence 2, Application US/09511477  
; Patent No. 6337202  
; GENERAL INFORMATION:  
; APPLICANT: JACOBSON, Myron K.  
; APPLICANT: JACOBSON, Elaine L.  
; APPLICANT: AM, Jean-Christophe  
; APPLICANT: LIN, Winston  
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY(ADP-RIBOSE) GLYCOHYDROLASE (PARG) EN  
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIVE  
; TITLE OF INVENTION: THEREWITH  
; FILE REFERENCE: NIAD 201  
; CURRENT APPLICATION NUMBER: US/09/511,477  
; CURRENT FILING DATE: 2000-02-23  
; PRIOR APPLICATION NUMBER: 09/302,812  
; PRIOR FILING DATE: 1999-04-30  
; NUMBER OF SEQ ID NOS: 38  
; SEQ ID NO 2  
; LENGTH: 977  
; TYPE: PRT  
; ORGANISM: Bos taurus  
; FEATURE:  
US-09-511-477-2

Query Match 100.0%; Score 53; DB 4; Length 977;  
Best Local Similarity 100.0%; Pred. No. 0.037; Mismatches 0; Indels 0; Gaps 0;  
Matches 11; Conservative 0;

QY 1 FLINPELIVSR 11  
DB 751 FLINPELIVSR 761

Db 760 FLINPELIVSR 770

RESULT 9  
US-09-511-507-2  
; Sequence 2, Application US/09511507  
; Patent No. 639543  
; GENERAL INFORMATION:  
; APPLICANT: JACOBSON, Myron K.  
; APPLICANT: JACOBSON, Elaine L.  
; APPLICANT: AM, Jean-Christophe  
; APPLICANT: LIN, Winston  
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY(ADP-RIBOSE) GLYCOHYDROLASE (PARG) ENZ  
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIVE  
; TITLE OF INVENTION: THEREWITH  
; FILE REFERENCE: NIAD 201  
; CURRENT APPLICATION NUMBER: US/09/511,507  
; CURRENT FILING DATE: 2000-02-23  
; PRIOR APPLICATION NUMBER: 09/302,812  
; PRIOR FILING DATE: 1999-04-30  
; NUMBER OF SEQ ID NOS: 38  
; SEQ ID NO 2  
; LENGTH: 977  
; TYPE: PRT  
; ORGANISM: Bos taurus  
; FEATURE:  
US-09-511-507-2

Query Match 100.0%; Score 53; DB 4; Length 977;  
Best Local Similarity 100.0%; Pred. No. 0.037; Mismatches 0; Indels 0; Gaps 0;  
Matches 11; Conservative 0;

QY 1 FLINPELIVSR 11  
DB 760 FLINPELIVSR 770

RESULT 10  
US-09-302-812-4  
; Sequence 4, Application US/09302812B  
; Patent No. 6333148  
; GENERAL INFORMATION:  
; APPLICANT: JACOBSON, Myron K.  
; APPLICANT: JACOBSON, Elaine L.  
; APPLICANT: AM, Jean-Christophe  
; APPLICANT: LIN, Winston  
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY(ADP-RIBOSE) GLYCOHYDROLASE (PARG) ENZ  
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIVE  
; TITLE OF INVENTION: THEREWITH  
; FILE REFERENCE: NIAD 201  
; CURRENT APPLICATION NUMBER: US/09/302,812B  
; CURRENT FILING DATE: 1999-04-30  
; EARLIER APPLICATION NUMBER: 60/083,768  
; EARLIER FILING DATE: 1998-05-01  
; NUMBER OF SEQ ID NOS: 38  
; SEQ ID NO 4  
; LENGTH: 976  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
US-09-302-812-4

Query Match 98.1%; Score 52; DB 4; Length 976;  
Best Local Similarity 90.9%; Pred. No. 0.058; Mismatches 1; Indels 0; Gaps 0;  
Matches 10; Conservative 1;

QY 1 FLINPELIVSR 11  
DB 759 FLINPELIVSR 769

RESULT 11  
US-09-511-477-4

; Sequence 4, Application US/09511477  
; Patent No. 6337202  
; GENERAL INFORMATION:  
; APPLICANT: JACOBSON, Myron K.  
; APPLICANT: JACOBSON, Elaine L.  
; APPLICANT: AME, Jean-Christophe  
; APPLICANT: LIN, Winston  
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE (PARG) EN  
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV  
; TITLE OF INVENTION: THEREWITH  
; FILE REFERENCE: NIAD 201  
; CURRENT APPLICATION NUMBER: US/09/511.477  
; CURRENT FILING DATE: 2000-02-23  
; PRIOR APPLICATION NUMBER: 09/302,812  
; PRIOR FILING DATE: 1999-04-30  
; NUMBER OF SEQ ID NOS: 38  
; SEQ ID NO 4  
; LENGTH: 976  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
US-09-511-477-4  
Query Match 98.1%; Score 52; DB 4; Length 976;  
Best Local Similarity 90.9%; Pred. No. 0.058;  
Matches 10; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
QY 1 FLINPELIVSR 11  
Db 759 FLINPELIISR 769  
RESULT 12  
US-09-511-507-4  
; Sequence 4, Application US/09511507  
; Patent No. 6395543  
; GENERAL INFORMATION:  
; APPLICANT: JACOBSON, Myron K.  
; APPLICANT: JACOBSON, Elaine L.  
; APPLICANT: AME, Jean-Christophe  
; APPLICANT: LIN, Winston  
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE (PARG) EN  
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV  
; TITLE OF INVENTION: THEREWITH  
; FILE REFERENCE: NIAD 201  
; CURRENT APPLICATION NUMBER: US/09/511.507  
; CURRENT FILING DATE: 2000-02-23  
; PRIOR APPLICATION NUMBER: 09/302,812  
; PRIOR FILING DATE: 1999-04-30  
; NUMBER OF SEQ ID NOS: 38  
; SEQ ID NO 4  
; LENGTH: 976  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
US-09-511-507-4  
Query Match 98.1%; Score 52; DB 4; Length 976;  
Best Local Similarity 90.9%; Pred. No. 0.058;  
Matches 10; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
QY 1 FLINPELIVSR 11  
Db 759 FLINPELIISR 769  
RESULT 13  
US-09-198-452A-694  
; Sequence 694, Application US/09198452A  
; Patent No. 6559294  
; GENERAL INFORMATION:  
; APPLICANT: Griffais, R.  
; TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fragments

; TITLE OF INVENTION: thereof and uses thereof, in particular for the diagnosis, prevent  
; TITLE OF INVENTION: and treatment of infection  
; FILE REFERENCE: 9710-003-999  
; CURRENT APPLICATION NUMBER: US/09/198,452A  
; CURRENT FILING DATE: 1998-11-24  
; NUMBER OF SEQ ID NOS: 6849  
; SEQ ID NO 694  
; LENGTH: 252  
; TYPE: PRT  
; ORGANISM: Chlamydia pneumoniae  
US-09-198-452A-694  
Query Match 77.4%; Score 41; DB 4; Length 252;  
Best Local Similarity 80.0%; Pred. No. 1.9;  
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
QY 1 FLINPELIVS 10  
Db 48 FLINPERVVS 57  
RESULT 14  
US-08-026-138E-4  
; Sequence 4, Application US/08026138E  
; Patent No. 5502166  
; GENERAL INFORMATION:  
; APPLICANT: Masayoshi MISHINA  
; TITLE OF INVENTION: NOVEL PROTEINS AND GENES CODING THE SAME  
; NUMBER OF SEQUENCES: 19  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Nishiohata Residence 1-107  
; STREET: 5214, Nishiohata-machi  
; CITY: Niigata-shi  
; STATE: Niigata-ken  
; COUNTRY: JAPAN  
; ZIP: 951  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.50 inch, 1.44 MB storage  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: MS-DOS v.5  
; SOFTWARE: Word Perfect 5.1  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: JP 39563/1992  
; FILING DATE: 26-FEB-1992  
; APPLICATION NUMBER: US/08/026,138E  
; FILING DATE: 26-FEB-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: JP 39563/1992  
; FILING DATE: 26-FEB-1992  
; APPLICATION NUMBER: JP 173155/1992  
; FILING DATE: 30-JUN-1992  
; APPLICATION NUMBER: JP 215017/1992  
; FILING DATE: 12-AUG-1992  
; APPLICATION NUMBER: JP 303878/1992  
; FILING DATE: 13-NOV-1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Hamburg, C.Bruce  
; REGISTRATION NUMBER: 22,389  
; REFERENCE/DOCKET NUMBER: F-4551  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212) 986-2340  
; TELEFAX: (212) 953-7733  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1323 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single strand  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; ORIGINAL SOURCE:  
; ORGANISM: mouse  
; TISSUE TYPE: brain  
; PUBLICATION INFORMATION:  
; AUTHORS: Masayoshi MISHINA  
; TITLE: NOVEL PROTEINS AND GENES CODING THE SAME

Don't take out anything

RELEVANT RESIDUES IN SEQ ID NO: 4: FROM 1 to 1323  
US-08-026-138E-4

Query Match 69.8%; Score 37; DB 1; Length 1323;  
Best Local Similarity 66.7%; Pred. No. 88;  
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 FLINPELIV 9  
Db 378 FLVNPESLV 386

RESULT 15  
US-08-231-193A-58  
; Sequence 58, Application US/08231193A  
; Patent No. 5849895  
; GENERAL INFORMATION:  
; APPLICANT: Daggett, Lorrie P.  
; APPLICANT: Ellis, Steven B.  
; APPLICANT: Liaw, Chen W.  
; APPLICANT: Lu, Chin-Chun  
; TITLE OF INVENTION: HUMAN N-METHYL-D-ASPARTATE RECEPTOR  
; SUBUNITS, NUCLEIC ACIDS ENCODING SAME AND USES THEREFOR  
; NUMBER OF SEQUENCES: 63  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Brown, Martin, Haller & McClain  
; STREET: 1660 Union Street  
; CITY: San Diego  
; STATE: CA  
; COUNTRY: U.S.A.  
; ZIP: 92101-2926  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: IBM PC compatible  
; SOFTWARE: Patent in Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/231,193A  
; FILING DATE: 20-APR-1994  
; CLASSIFICATION: 536  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/052,459  
; FILING DATE: 20-APR-1993  
; CLASSIFICATION: 536  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Seidman, Stephanie  
; REGISTRATION NUMBER: 33,779  
; REFERENCE/DOCKET NUMBER: 6362-9383  
; TELEPHONE: 619-238-0999  
; TELEFAX: 619-238-0062  
; INFORMATION FOR SEQ ID NO: 58:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1336 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-231-193A-58

Query Match 69.8%; Score 37; DB 2; Length 1336;  
Best Local Similarity 66.7%; Pred. No. 89;  
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 FLINPELIV 9  
Db 381 FLVNPESLV 389

Search completed: May 26, 2004, 18:49:32  
Job time : 1.35248 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 26, 2004, 18:46:21 ; Search time 0.917775 Seconds  
(without alignments)  
3343.018 Million cell updates/sec

Title: US-09-302-812-13  
Perfect score: 53  
Sequence: 1 FLINPELIVSR 11

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1149313 seqs, 278921704 residues

Total number of hits satisfying chosen parameters: 1149313

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications AA.\*

- 1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*
- 2: /cgn2\_6/ptodata/1/pubpaa/FCT\_NEW\_PUB.pep.\*
- 3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep.\*
- 4: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*
- 5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep.\*
- 6: /cgn2\_6/ptodata/1/pubpaa/PCTUS\_PUBCOMB.pep.\*
- 7: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep.\*
- 8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep.\*
- 9: /cgn2\_6/ptodata/1/pubpaa/US09A\_PUBCOMB.pep.\*
- 10: /cgn2\_6/ptodata/1/pubpaa/US09B\_PUBCOMB.pep.\*
- 11: /cgn2\_6/ptodata/1/pubpaa/US09C\_PUBCOMB.pep.\*
- 12: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep.\*
- 13: /cgn2\_6/ptodata/1/pubpaa/US10A\_PUBCOMB.pep.\*
- 14: /cgn2\_6/ptodata/1/pubpaa/US10B\_PUBCOMB.pep.\*
- 15: /cgn2\_6/ptodata/1/pubpaa/US10C\_PUBCOMB.pep.\*
- 16: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep.\*
- 17: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep.\*
- 18: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	53	100.0	11	9	US-09-973-451-13
2	53	100.0	968	9	US-09-973-451-6
3	53	100.0	977	9	US-09-973-451-2
4	52	98.1	976	9	US-09-973-451-4
5	42	79.2	119	12	US-10-424-599-168770
6	42	79.2	546	12	US-10-425-114-60000
7	41	77.4	249	12	US-10-282-122A-54709
8	41	77.4	252	15	US-10-289-762-694
9	38	71.7	59	12	US-10-424-599-215648
10	37	69.8	1336	9	US-09-945-901-58
11	37	69.8	1336	13	US-10-007-747-58
12	37	69.8	1336	14	US-10-038-937-58
13	36	67.9	107	12	US-10-424-599-183741
14	35	66.0	449	12	US-10-282-122A-73824
15	35	66.0	527	15	US-10-369-493-1735

16	35	66.0	621	14	US-10-156-761-10025	Sequence 10025, A
17	35	66.0	820	14	US-10-174-677-30	Sequence 30, Appl
18	35	66.0	828	14	US-10-174-677-90	Sequence 90, Appl
19	35	66.0	932	14	US-10-174-677-89	Sequence 89, Appl
20	35	66.0	932	14	US-10-174-677-91	Sequence 91, Appl
21	34	64.2	21	16	US-10-663-896-17	Sequence 17, Appl
22	34	64.2	39	12	US-10-424-599-231144	Sequence 231144, A
23	34	64.2	72	12	US-10-424-599-176489	Sequence 176489, A
24	34	64.2	86	12	US-10-424-599-183835	Sequence 183835, A
25	34	64.2	98	12	US-10-424-599-183834	Sequence 183834, A
26	34	64.2	100	12	US-10-424-599-162739	Sequence 162739, A
27	34	64.2	138	10	US-09-896-580A-3	Sequence 8, Appl
28	34	64.2	169	10	US-09-896-580A-8	Sequence 3, Appl
29	34	64.2	211	12	US-10-282-122A-47127	Sequence 47127, A
30	34	64.2	275	12	US-10-424-599-169986	Sequence 169986, A
31	34	64.2	307	12	US-10-425-114-58003	Sequence 58003, A
32	34	64.2	407	12	US-10-282-122A-63173	Sequence 63173, A
33	34	64.2	408	15	US-10-369-493-3561	Sequence 3561, Ap
34	34	64.2	449	15	US-10-369-493-20063	Sequence 20063, A
35	34	64.2	451	12	US-10-282-122A-72152	Sequence 72152, A
36	34	64.2	651	15	US-10-369-493-21499	Sequence 21499, A
37	34	64.2	761	12	US-10-114-270-108	Sequence 108, App
38	34	64.2	1065	9	US-09-771-161A-239	Sequence 239, App
39	33	62.3	58	12	US-10-424-599-211138	Sequence 211138, A
40	33	62.3	59	9	US-09-864-761-38589	Sequence 38589, A
41	33	62.3	241	12	US-10-424-599-255332	Sequence 255332, A
42	33	62.3	269	12	US-10-282-122A-71472	Sequence 71472, A
43	33	62.3	295	12	US-10-282-122A-53386	Sequence 53386, A
44	33	62.3	428	12	US-10-282-122A-54992	Sequence 54992, A
45	33	62.3	428	15	US-10-312-273-93	Sequence 93, Appl

## ALIGNMENTS

### RESULT 1

US-09-973-451-13  
; Sequence 13, Application US/09973451  
; Patent No. US20020132328A1  
; GENERAL INFORMATION:  
; APPLICANT: JACOBSON, Myron K.  
; APPLICANT: JACOBSON, Elaine L.  
; APPLICANT: AM, Jean-Christophe  
; APPLICANT: LIN, Winston  
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE  
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIVE  
; TITLE OF INVENTION: THEREWITH  
; FILE REFERENCE: NIAD 201  
; CURRENT APPLICATION NUMBER: US/09/973,451  
; CURRENT FILING DATE: 2001-10-09  
; PRIOR APPLICATION NUMBER: US/09/302,812  
; PRIOR FILING DATE: 1999-04-30  
; PRIOR APPLICATION NUMBER: 60/083,768  
; PRIOR FILING DATE: 1998-05-01  
; NUMBER OF SEQ ID NOS: 38  
; SEQ ID NO 13  
; LENGTH: 11  
; TYPE: PRT  
; ORGANISM: Bos taurus  
; FEATURE:  
US-09-973-451-13

Query Match 100.0%; Score 53; DB 9; Length 11;  
Best Local Similarity 100.0%; Pred. No. 0.0025;  
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 FLINPELIVSR 11

Db 1 FLINPELIVSR 11

### RESULT 2



US-09-973-451-6  
; Sequence 6, Application US/09973451  
; Patent No. US20020132328A1  
; GENERAL INFORMATION:  
; APPLICANT: JACOBSON, Myron K.  
; APPLICANT: JACOBSON, Elaine L.  
; APPLICANT: AM, Jean-Christophe  
; APPLICANT: LIN, Winston  
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE  
; TITLE OF INVENTION: (PARG) ENZYMES,  
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIVE  
; TITLE OF INVENTION: THEREWITH  
; FILE REFERENCE: NIA2 201  
; CURRENT APPLICATION NUMBER: US/09/973,451  
; CURRENT FILING DATE: 2001-10-09  
; PRIOR APPLICATION NUMBER: US/09/302,812  
; PRIOR FILING DATE: 1999-04-30  
; PRIOR APPLICATION NUMBER: 60/083,768  
; PRIOR FILING DATE: 1998-05-01  
; NUMBER OF SEQ ID NOS: 38  
; SEQ ID NO 6  
; LENGTH: 968  
; TYPE: PRT  
; ORGANISM: Mus musculus  
; FEATURE:  
US-09-973-451-6

Query Match 100.0%; Score 53; DB 9; Length 968;  
Best Local Similarity 100.0%; Pred. No. 0.35;  
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 FLINPELIVSR 11  
|||  
Db 751 FLINPELIVSR 761

RESULT 3  
US-09-973-451-2  
; Sequence 2, Application US/09973451  
; Patent No. US20020132328A1  
; GENERAL INFORMATION:  
; APPLICANT: JACOBSON, Myron K.  
; APPLICANT: JACOBSON, Elaine L.  
; APPLICANT: AM, Jean-Christophe  
; APPLICANT: LIN, Winston  
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE  
; TITLE OF INVENTION: (PARG) ENZYMES,  
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIVE  
; TITLE OF INVENTION: THEREWITH  
; FILE REFERENCE: NIA2 201  
; CURRENT APPLICATION NUMBER: US/09/973,451  
; CURRENT FILING DATE: 2001-10-09  
; PRIOR APPLICATION NUMBER: US/09/302,812  
; PRIOR FILING DATE: 1999-04-30  
; PRIOR APPLICATION NUMBER: 60/083,768  
; PRIOR FILING DATE: 1998-05-01  
; NUMBER OF SEQ ID NOS: 38  
; SEQ ID NO 2  
; LENGTH: 977  
; TYPE: PRT  
; ORGANISM: Bos taurus  
; FEATURE:  
US-09-973-451-2

Query Match 100.0%; Score 53; DB 9; Length 977;  
Best Local Similarity 100.0%; Pred. No. 0.35;  
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 FLINPELIVSR 11  
|||  
Db 760 FLINPELIVSR 770

RESULT 4  
US-09-973-451-4  
; Sequence 4, Application US/09973451  
; Patent No. US20020132328A1  
; GENERAL INFORMATION:  
; APPLICANT: JACOBSON, Myron K.  
; APPLICANT: JACOBSON, Elaine L.  
; APPLICANT: AM, Jean-Christophe  
; APPLICANT: LIN, Winston  
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE  
; TITLE OF INVENTION: (PARG) ENZYMES,  
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIVE  
; TITLE OF INVENTION: THEREWITH  
; FILE REFERENCE: NIA2 201  
; CURRENT APPLICATION NUMBER: US/09/973,451  
; CURRENT FILING DATE: 2001-10-09  
; PRIOR APPLICATION NUMBER: US/09/302,812  
; PRIOR FILING DATE: 1999-04-30  
; PRIOR APPLICATION NUMBER: 60/083,768  
; PRIOR FILING DATE: 1998-05-01  
; NUMBER OF SEQ ID NOS: 38  
; SEQ ID NO 4  
; LENGTH: 976  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
US-09-973-451-4

Query Match 98.1%; Score 52; DB 9; Length 976;  
Best Local Similarity 90.9%; Pred. No. 0.54;  
Matches 10; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 FLINPELIVSR 11  
|||  
Db 759 FLINPELIVSR 769

RESULT 5  
US-10-424-599-168770  
; Sequence 168770, Application US/10424599  
; Publication No. US20040031072A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa Thomas J  
; APPLICANT: Kovalic David K  
; APPLICANT: Zhou Yihua  
; APPLICANT: Cao Yongwei  
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 38-21(53223)B  
; CURRENT APPLICATION NUMBER: US/10/424,599  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 285684  
; SEQ ID NO 168770  
; LENGTH: 119  
; TYPE: PRT  
; ORGANISM: Glycine max  
; FEATURE:  
; NAME/KEY: unsure  
; LOCATION: (1)..(119)  
; OTHER INFORMATION: unsure at all Xaa locations  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT3847\_123414C.1.pep  
US-10-424-599-168770

Query Match 79.2%; Score 42; DB 12; Length 119;  
Best Local Similarity 63.6%; Pred. No. 4;  
Matches 7; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Qy 1 FLINPELIVSR 11  
|||  
Db 71 FLVQPELVNVR 81

```
RESULT 6
US-10-425-114-60000
; Sequence 60000, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(5313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 60000
; LENGTH: 546
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB3689-262-B12_FLI.pep
US-10-425-114-60000

Query Match 79.2%; Score 42; DB 12; Length 546;
Best Local Similarity 88.9%; Pred. No. 22;
Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 FLINPELV 9
DB 319 FMINPELV 327

RESULT 7
US-10-282-122A-54709
; Sequence 54709, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELIIRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09

US-09-302-812-13.rapb
```

```
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 54709
; LENGTH: 249
; TYPE: PRT
; ORGANISM: Chlamydia pneumoniae
US-10-282-122A-54709

Query Match 77.4%; Score 41; DB 12; Length 249;
Best Local Similarity 80.0%; Pred. No. 14;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 FLINPELV 10
DB 45 FLINPERVVS 54

RESULT 8
US-10-289-762-694
; Sequence 694, Application US/10289762
; Publication No. US20040006218A1
; GENERAL INFORMATION:
; APPLICANT: Griffais, R.
; TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fragments
; TITLE OF INVENTION: thereof and uses thereof, in particular for the diagnosis, prevent
; TITLE OF INVENTION: and treatment of infection
; FILE REFERENCE: 9710-003-999
; CURRENT APPLICATION NUMBER: US/10/289,762
; CURRENT FILING DATE: 2003-03-27
; NUMBER OF SEQ ID NOS: 6849
; SEQ ID NO 694
; LENGTH: 252
; TYPE: PRT
; ORGANISM: Chlamydia pneumoniae
US-10-289-762-694

Query Match 77.4%; Score 41; DB 15; Length 252;
Best Local Similarity 80.0%; Pred. No. 14;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 FLINPELV 10
DB 48 FLINPERVVS 57

RESULT 9
US-10-424-599-215648
; Sequence 215648, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J
; APPLICANT: Kovalic, David K
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(5323)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 215648
; LENGTH: 59
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(59)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_36759C.1.pap
```

US-10-424-599-215648

Query Match 71.7%; Score 38; DB 12; Length 59;  
Best Local Similarity 45.5%; Pred. No. 11;  
Matches 5; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

QY 1 FLINPELVSR 11  
||: ||: ||: ||:  
Db 22 FLLEPENVIIR 32

RESULT 10

US-09-945-901-58  
; Sequence 58, Application US/09945901  
; Patent No. US20020161215A1

GENERAL INFORMATION:

APPLICANT: Daggett, Lorrie P.  
Ellis, Steven B.  
Liaw, Chen W.  
Lu, Chin-Chun

TITLE OF INVENTION: HUMAN N-METHYL-D-ASPARTATE RECEPTOR

SUBUNITS, NUCLEIC ACIDS ENCODING SAME AND USES THEREFOR

NUMBER OF SEQUENCES: 63

CORRESPONDENCE ADDRESS:

ADDRESSEE: Heller Ehrman White & McAuliffe  
STREET: 4250 Executive Square, 7th Floor  
CITY: La Jolla  
STATE: CA  
COUNTRY: U.S.A.

ZIP: 92037

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/945,901

FILING DATE: 24-Jan-2001

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/940,035

FILING DATE: <Unknown>

APPLICATION NUMBER: US 08/052,449

FILING DATE: 20-APR-1993

ATTORNEY/AGENT INFORMATION:

NAME: Seidman, Stephanie

REGISTRATION NUMBER: 33,779

REFERENCE/DOCKET NUMBER: 6362-9383E

TELECOMMUNICATION INFORMATION:

TELEPHONE: 619-238-0999

TELEFAX: 619-238-0062

INFORMATION FOR SEQ ID NO: 58:

SEQUENCE CHARACTERISTICS:

LENGTH: 1336 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

SEQUENCE DESCRIPTION: SEQ ID NO: 58:

US-09-945-901-58

Query Match 69.8%; Score 37; DB 9; Length 1336;  
Best Local Similarity 66.7%; Pred. No. 5.1e+02;  
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 FLINPELVSR 9  
||: ||: ||: ||:  
Db 381 FLVNPSLVV 389

RESULT 11

US-10-007-747-58

; Sequence 58, Application US/10007747

; Publication No. US20020161193A1

GENERAL INFORMATION:

APPLICANT: Daggett, Lorrie P.  
Ellis, Steven B.  
Liaw, Chen W.  
Lu, Chin-Chun

TITLE OF INVENTION: HUMAN N-METHYL-D-ASPARTATE RECEPTOR

SUBUNITS, NUCLEIC ACIDS ENCODING SAME AND USES THEREFOR

NUMBER OF SEQUENCES: 63

CORRESPONDENCE ADDRESS:

ADDRESSEE: Heller Ehrman White & McAuliffe  
STREET: 4250 Executive Square, 7th Floor  
CITY: La Jolla  
STATE: CA  
COUNTRY: USA

ZIP: 92037

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/10/007,747

FILING DATE: 07-Dec-2001

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/09/648,797

FILING DATE: 28-Aug-2000

APPLICATION NUMBER: US/08/940,086A

FILING DATE: 29-SEPT-97

APPLICATION NUMBER: US 08/231,193

FILING DATE: 20-APR-1994

APPLICATION NUMBER: US 08/052,449

FILING DATE: 20-APR-1993

ATTORNEY/AGENT INFORMATION:

NAME: Seidman, Stephanie

REGISTRATION NUMBER: 33,779

REFERENCE/DOCKET NUMBER: 24735-9383C

TELECOMMUNICATION INFORMATION:

TELEPHONE: (619) 450-8400

TELEFAX: (619) 450-8499

INFORMATION FOR SEQ ID NO: 58:

SEQUENCE CHARACTERISTICS:

LENGTH: 1336 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

SEQUENCE DESCRIPTION: SEQ ID NO: 58:

US-10-007-747-58

Query Match 69.8%; Score 37; DB 13; Length 1336;

Best Local Similarity 66.7%; Pred. No. 5.1e+02;

Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 FLINPELVSR 9  
||: ||: ||: ||:  
Db 381 FLVNPSLVV 389

RESULT 12

US-10-038-937-58

; Sequence 58, Application US/10038937

; Publication No. US20030013866A1

GENERAL INFORMATION:

APPLICANT: Daggett, Lorrie P.

Lu, Chin-Chun

TITLE OF INVENTION: HUMAN N-METHYL-D-ASPARTATE RECEPTOR

SUBUNITS, NUCLEIC ACIDS ENCODING SAME AND USES THEREFOR

NUMBER OF SEQUENCES: 63

CORRESPONDENCE ADDRESS:

ADDRESSEE: Heller Ehrman White & McAuliffe  
STREET: 4250 Executive Square, 7th Floor  
CITY: La Jolla  
STATE: CA  
COUNTRY: U.S.A.

ZIP: 92037  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA: US/10/038,937  
APPLICATION NUMBER: US/10/038,937  
FILING DATE: 18-Apr-2002  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/935,105  
FILING DATE: 29-SEP-97  
APPLICATION NUMBER: US 08/231,193  
FILING DATE: 20-APR-1994  
APPLICATION NUMBER: US 08/052,449  
FILING DATE: 20-APR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Seidman, Stephanie  
REGISTRATION NUMBER: 33,779  
REFERENCE/DOCKET NUMBER: 6362-9383D  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 619-238-0999  
TELEFAX: 619-238-0062  
INFORMATION FOR SEQ ID NO: 58:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1336 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
SEQUENCE DESCRIPTION: SEQ ID NO: 58:  
US-10-038-937-58

Query Match 69.8%; Score 37; DB 14; Length 1336;  
Best Local Similarity 66.7%; Pred. No. 5.1e+02;  
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 FLINPELIV 9  
DB 381 FLVNPFLV 389  
RESULT 13  
US-10-424-599-183741  
; Sequence 183741, Application US/10424599  
; Publication No. US20040031072A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa, Thomas J  
; APPLICANT: Kovalic, David K  
; APPLICANT: Zhou, Yihua  
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 38-21(53223)B  
; CURRENT APPLICATION NUMBER: US/10/424,599  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 285684  
; SEQ ID NO 183741  
; LENGTH: 107  
; TYPE: PRT  
; ORGANISM: Glycine max  
; FEATURE:  
; NAME/KEY: unsure  
; LOCATION: (1)..(107)  
; OTHER INFORMATION: unsure at all Xaa locations  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT3847\_136932C.1.pep  
US-10-424-599-183741  
Query Match 67.9%; Score 36; DB 12; Length 107;  
Best Local Similarity 60.0%; Pred. No. 48;  
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 FLINPELIV 10  
DB 94 FWINPTLIIN 103

RESULT 14  
US-10-282-122A-73824  
; Sequence 73824, Application US/10282122A  
; Publication No. US20040029129A1  
; GENERAL INFORMATION:  
; APPLICANT: Wang, Liangsu  
; APPLICANT: Zamudio, Carlos  
; APPLICANT: Malone, Cheryl  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Kari  
; APPLICANT: Zyskind, Judith  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John  
; APPLICANT: Carr, Grant  
; APPLICANT: Yamamoto, Robert  
; APPLICANT: Forsyth, R.  
; APPLICANT: Xu, H.  
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms  
; FILE REFERENCE: ELITRA.034A  
; CURRENT APPLICATION NUMBER: US/10/282,122A  
; CURRENT FILING DATE: 2003-02-20  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/230,335  
; PRIOR FILING DATE: 2000-09-06  
; PRIOR APPLICATION NUMBER: 60/230,347  
; PRIOR FILING DATE: 2000-09-09  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/267,636  
; PRIOR FILING DATE: 2001-02-09  
; PRIOR APPLICATION NUMBER: 60/269,308  
; PRIOR FILING DATE: 2001-02-16  
; Remaining Prior Application data removed - See File Wrapper or PALM.  
; NUMBER OF SEQ ID NOS: 78614  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 73824  
; LENGTH: 449  
; TYPE: PRT  
; ORGANISM: Streptococcus pneumoniae  
US-10-282-122A-73824  
Query Match 66.0%; Score 35; DB 12; Length 449;  
Best Local Similarity 50.0%; Pred. No. 3.6e+02;  
Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 2 LINPELIVSR 11  
DB 119 IVNPPELLAAR 128  
RESULT 15  
US-10-369-493-1735  
; Sequence 1735, Application US/10369493  
; Publication No. US20030233675A1  
; GENERAL INFORMATION:  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Goldman, Barry S.

```

; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 1735
; LENGTH: 527
; TYPE: PRT
; ORGANISM: Saccharomyces cerevisiae
; US-10-369-493-1735

```

```

Query Match      66.0%; Score 35; DB 15; Length 527;
Best Local Similarity 63.6%; Pred. No. 4.3e+02;
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

```

```

QY      1 FLINPELIVSR 11
Db      309 FMISPRILIVIR 319

```

Search completed: May 26, 2004, 19:19:02  
 Job time : 1.91778 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 26, 2004, 17:20:49 ; Search time 0.176239 Seconds  
(without alignments)  
3249.964 Million cell updates/sec

Title: US-09-302-812-13  
Perfect score: 53  
Sequence: 1 FLINPELIVSR 11

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_42.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	40	75.5	316	1	SRG4 CAEEL	P54126 caenorhabdi
2	39	73.6	314	1	NOD1 RHILP	P23718 rhizobium 1
3	38	71.7	150	1	DEF CLOTE	Q895g2 clostridium
4	37	69.8	295	1	Y198 CLOPE	P26832 clostridium
5	37	69.8	1323	1	NME4 MOUSE	Q03391 mus musculus
6	37	69.8	1323	1	NME4 MOUSE	Q62845 rattus norv
7	37	69.8	1336	1	NME4 HUMAN	O15399 homo sapien
8	36	67.9	185	1	DEF2 NITEU	Q82tc8 nitrosomona
9	36	67.9	320	1	FLIG BUCBP	Q89az9 buchnera ap
10	36	67.9	714	1	ZW10 DROGR	O44219 drosophila
11	35	66.0	169	1	DEF1 VIBCH	Q9kvv3 vibrio chol
12	35	66.0	175	1	DEF1 RICCN	Q924z1 rickettsia
13	35	66.0	175	1	DEF RICEP	Q92dv8 rickettsia
14	35	66.0	527	1	IF2G YEAST	P32481 saccharomyc
15	35	66.0	839	1	Y422 MYCPN	P75175 mycoplasma
16	35	66.0	932	1	CDG8 HUMAN	Q9v5g5 homo sapien
17	35	66.0	932	1	CDG9 HUMAN	Q9v5g4 homo sapien
18	34	64.2	112	1	YFTA_ECOLI	P11285 escherichia
19	34	64.2	169	1	DEF2 HABIN	P47866 haemophilus
20	34	64.2	201	1	DEF2 PROMM	Q47v5f9 prochloroco
21	34	64.2	211	1	KAD BORBU	O51378 borrelia bu
22	34	64.2	215	1	TDX1 SULME	Q55060 sulfolobus
23	34	64.2	245	1	YNFI_RHOCA	P17435 rhodobacter
24	34	64.2	246	1	MCT1_MERUN	P50340 meriones un
25	34	64.2	270	1	PNK1_LACLA	Q9c1j4 lactococcus
26	34	64.2	285	1	Y011 MYCPN	P75098 mycoplasma
27	34	64.2	651	1	Y942 METJA	Q58352 methanococc
28	34	64.2	726	1	YBID SCHPO	P87178 schizosacch
29	34	64.2	1065	1	KDGI_HUMAN	Q75912 homo sapien
30	33	62.3	130	1	NSUM STRFN	Q35832 strongyloe
31	33	62.3	170	1	DEF1 VIBVU	Q8dcd3 vibrio vuln
32	33	62.3	170	1	DEF PASMU	P57948 pasteurella
33	33	62.3	172	1	DEF1_VIBPA	Q87kds vibrio para

34	33	62.3	202	1	DEF2 VIBVY	Q7mgk6 vibrio vuln
35	33	62.3	251	1	Y545 METUA	Q57965 methanococc
36	33	62.3	305	1	EFTS_BRUME	Q8vhh5 brucella me
37	33	62.3	313	1	NPR_BACSU	P39790 bacillus su
38	33	62.3	314	1	NOD2_RHITR	P32008 rhizobium t
39	33	62.3	326	1	ACCD SYN33	Q57417 synechocyst
40	33	62.3	428	1	ENO CHLPN	Q927a6 chlamydia p
41	33	62.3	479	1	ZW10 DROPS	O44218 drosophila
42	33	62.3	709	1	ENT2_FOWPV	Q9j562 fowlpox vir
43	33	62.3	751	1	PSNA_CHLVU	P56341 chlorella v
44	33	62.3	800	1	GUN_BACSI	P08564 bacillus sp
45	33	62.3	875	1	TRAC_ECOLI	P18004 escherichia

## ALIGNMENTS

RESULT 1  
SRG4 CAEEL STANDARD; PRT; 316 AA.  
AC P54126; Q95ZP2;  
DT 01-OCT-1996 (Rel. 34, Created)  
DT 10-OCT-2003 (Rel. 42, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Serpentine receptor class gamma 4 (Srg-4 protein).  
GN SRG-4 OR T12A2.12.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
OC Rhabditidae; Peloderinae; Caenorhabditis.  
OX NCBI\_TaxID=6239;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Bristol N2;  
RA Latreille P.;  
RL Submitted (NOV-1994) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP REVISIONS.  
RA Waterston R.;  
RL Submitted (JAN-2003) to the EMBL/GenBank/DBJ databases.  
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Probable).  
CC -!- SIMILARITY: Belongs to the C.elegans receptor-like protein srg family.

-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----

EMBL; U13019; AAK84568.2; --  
PIR; T15557; T15557.  
WormPep; T12A2.12; CB33181.  
InterPro; IPR000609; Srg.  
DR PRINTS; PR00698; TMPTOTINSRG.  
KW Transmembrane; Multigene family.  
FT TRANSMEM 21 41 POTENTIAL.  
FT TRANSMEM 52 72 POTENTIAL.  
FT TRANSMEM 88 108 POTENTIAL.  
FT TRANSMEM 140 160 POTENTIAL.  
FT TRANSMEM 188 208 POTENTIAL.  
FT TRANSMEM 229 249 POTENTIAL.  
FT TRANSMEM 259 279 POTENTIAL.  
SQ SEQUENCE 316 AA; 36716 MW; 05667CDFCC5E477D CRC64;

Query Match 75.5%; Score 40; DB 1; Length 316;  
Best Local Similarity 72.7%; Pred. No. 2.7;  
Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 FLINPELIVSR 11  
||| :|||  
DB 290 FLINFPVVS 300

```

SEQUENCE FROM N.A.
RP STRAIN=Massachusetts / E88;
RX MEDLINE=22457253; PubMed=12552129;
RA Brueggemann H., Baumer S., Fricke W.F., Wierzer A., Liesegang H.,
RA Decker I., Herzberg C., Martinez-Arias R., Merkl R., Henne A.,
RA Gottschalk G.;
RT "The genome sequence of Clostridium tetani, the causative agent of
RT tetanus disease.";
RL Proc. Natl. Acad. Sci. U.S.A. 100:1316-1321(2003).
CC -!- FUNCTION: Removes the formyl group from the N-terminal Met of
CC newly synthesized proteins. Requires at least a dipeptide for an
CC efficient rate of reaction. N-terminal L-methionine is a
CC prerequisite for activity but the enzyme has broad specificity at
CC other positions (By similarity).
CC -!- CATALYTIC ACTIVITY: Formyl-L-methionyl peptide + H(2)O = formate +
CC methionyl peptide.
CC -!- COFACTOR: Binds 1 iron(II) ion (By similarity).
CC -!- SIMILARITY: Belongs to the polypeptide deformylase family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AB015940; AAC35788.1; -.
DR HAMAP; MF_00163; -.
DR InterPro; IPR000181; Fmet deformylase.
DR Pfam; PF01327; Pep deformylase; 1.
DR PRINTS; PR01576; PDEFORMLASE.
DR PROSITE; PS00339; HTHLYSR.
KW Protein biosynthesis; Hydrolase; Iron; Complete proteome.
FT ACT_SITE 131 131 BY SIMILARITY.
FT METAL 88 88 IRON (BY SIMILARITY).
FT METAL 130 130 IRON (BY SIMILARITY).
FT METAL 134 134 IRON (BY SIMILARITY).
FT SEQUENCE 150 AA; 16924 MW; 2C0492E3688DC8E7 CRC64;
SQ
Query Match 71.7%; Score 38; DB 1; Length 150;
Best Local Similarity 87.5%; Pred. No. 3;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 FLINPELI 8
Db 69 FLINPELI 76

RESULT 4
Y188 CLOPE
ID Y188 CLOPE STANDARD; PRT; 295 AA.
AC P28832;
DT 01-AUG-1992 (Rel. 23, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hypothetical protein CPE0188.
GN CPE0188.
OS Clostridium perfringens
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1502;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=13 / Type A;
RX MEDLINE=21664373; PubMed=11792842;
RA Shimizu T., Ohtani K., Hirakawa H., Ohshima K., Yamashita A.,
RA Shiba T., Ogasawara N., Hattori M., Kuhara S., Hayashi H.;
RT "Complete genome sequence of Clostridium perfringens, an anaerobic
RT flesh-eater.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:996-1001(2002).
RN [2]
RP SEQUENCE OF 114-295 FROM N.A.
RC STRAIN=CPN50;

```

```

RESULT 2
NOD1 RHILP
ID NOD1 RHILP STANDARD; PRT; 314 AA.
AC P23718;
DT 01-NOV-1991 (Rel. 20, Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Modulation protein D I.
GN NODD1.
OS Rhizobium leguminosarum (biovar phaseoli).
OC Plasmid sym.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Rhizobiaceae; Rhizobium/Agrobacterium group; Rhizobium.
OX NCBI_TaxID=385;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=8002;
RX MEDLINE=91014692; PubMed=2215216;
RA Davis E.O., Johnston A.W.B.;
RT "Analysis of three nodD genes in Rhizobium leguminosarum biovar
RT phaseoli; nodD1 is preceded by noIE, a gene whose product is secreted
RT from the cytoplasm.";
RL Mol. Microbiol. 4:921-932(1990).
CC -!- FUNCTION: NODD REGULATES THE EXPRESSION OF THE NOD ABCFE GENES
CC WHICH ENCODE OTHER NODULATION PROTEINS. NODD IS ALSO A NEGATIVE
CC REGULATOR OF ITS OWN EXPRESSION. BINDS FLAVONOIDS AS INDUCERS.
CC -!- SIMILARITY: Contains 1 HTH LysR-type DNA-binding domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X54214; CAA38126.1; -.
DR PIR; S11787; S11787.
DR InterPro; IPR000847; HTH_LysR.
DR Pfam; PF00126; HTH 1; 1.
DR PRINTS; PR00039; HTHLYSR.
DR PROSITE; PS0931; HTH_LysR; 1.
KW Nodulation; Transcription regulation; DNA-binding; Activator;
KW Repressor; Plasmid.
FT DOMAIN 6 63 HTH LYSR-TYPE.
FT DNA_BIND 23 42 H-T-H MOTIF (BY SIMILARITY).
FT SEQUENCE 314 AA; 35476 MW; A17DBC9965DE32B6 CRC64;
SQ
Query Match 73.6%; Score 39; DB 1; Length 314;
Best Local Similarity 72.7%; Pred. No. 4.3;
Matches 8; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 FLINPELIVSR 11
Db 148 FLINPELIVSR 158

RESULT 3
DEF_CLOTE
ID DEF_CLOTE STANDARD; PRT; 150 AA.
AC Q895Q2;
DT 15-MAR-2004 (Rel. 43, Created)
DT 15-MAR-2004 (Rel. 43, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Peptide deformylase (EC 3.5.1.88) (PDF) (Polypeptide deformylase).
GN DEF OR CTC01219.
OS Clostridium tetani.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1513;
RN [1]

```

EX MEDLINE=94232189; PubMed=8177218;  
RA Canard B., Garnier T., Saint-Joanis B., Cole S.T.;  
RT "Molecular genetic analysis of the nahg gene encoding a hyaluronidase  
RI of Clostridium perfringens.";  
RL Mol. Genet. 243:215-224(1994).  
CC -1- SIMILARITY: BELONGS TO THE ROK (NAGC/XYLR) FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
CC or send an email to license@isb-sib.ch).  
CC -----  
CC DR EMBL; AP003185; BAB79894.1; -;  
CC DR EMBL; M81878; AAA23256.1; -;  
CC DR PIR; S43901; S43901.  
CC DR InterPro; IPR000600; ROK.  
CC DR Pfam; PF00480; ROK; 1. FALSE NEG.  
CC DR PROSITE; PS01125; ROK; Complete proteome.  
CC KW Hypothetical protein; Complete proteome.  
CC FT CONFLICT 114 122 ALGEFFGA -> NSRIIWGML (IN REF. 2).  
CC FT CONFLICT 168 168 H -> N (IN REF. 2).  
CC SQ SEQUENCE 295 AA; 31759 MW; 1CD289093872433 CRC64;  
  
Query Match 69.8%; Score 37; DB 1; Length 295;  
Best Local Similarity 66.7%; Pred. No. 9.9;  
Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 FLINPELIV 9  
DB 229 YLINPEVVV 237  
  
RESULT 5  
NME4\_MOUSE STANDARD; PRT; 1323 AA.  
AC Q03391;  
DT 01-JUN-1994 (Rel. 29, Created)  
DT 15-JUL-1999 (Rel. 38, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Glutamate [NMDA] receptor subunit epsilon 4 precursor (N-methyl  
DE D-aspartate receptor subtype 2D) (NR2D) (NMDAR2D).  
GN GRIN2D.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX TISSUE=Brain;  
RX MEDLINE=93050214; PubMed=1385220;  
RA Ikeda K., Nagasawa M., Mori H., Araki K., Sakimura K., Watanabe M.,  
RA Inoue Y., Mishina M.;  
RT "Cloning and expression of the epsilon 4 subunit of the NMDA receptor  
RT channel.";  
RL FEBS Lett. 313:34-38(1992).  
RN [2]  
RP REVISIONS.  
RA Ikeda K., Nagasawa M., Mori H., Araki K., Sakimura K., Watanabe M.,  
RA Inoue Y., Mishina M.;  
RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.  
CC -1- FUNCTION: NMDA receptor subtype of glutamate-gated ion channels  
CC possesses high calcium permeability and voltage-dependent  
CC sensitivity to magnesium and is mediated by glycine.  
CC -1- SUBUNIT: Heterodimer of an epsilon subunit and a zeta subunit.  
CC -1- SUBCELLULAR LOCATION: Integral membrane protein.  
CC -1- SIMILARITY: Belongs to the ligand-gated ion channel family.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See http://www.isb-sib.ch/announce/  
or send an email to license@isb-sib.ch).  
CC -----  
CC DR EMBL; D12822; BAA02254.1; -;  
CC DR HSSP; P19491; 1GR2.  
CC DR MGD; MGI:95823; Grin2d.  
CC DR InterPro; IPR001828; ANF\_receptor.  
CC DR InterPro; IPR001320; Ion\_glu\_receptor.  
CC DR InterPro; IPR001508; NMDA\_receptor.  
CC DR InterPro; IPR001311; SBP/glu\_receptor.  
CC Pfam; PF01094; ANF\_receptor; 1.  
CC Pfam; PF00060; lig\_chan; 1.  
CC PRINTS; PR00177; NMDARECEPTOR.  
CC SMART; SM00079; PBPE; 1.  
CC Receptor; Signal; Transmembrane; Postsynaptic membrane; Calcium;  
KW Ionic channel; Magnesium.  
KW SIGNAL 1 27  
FT CHAIN 28 1323  
FT DOMAIN 28 580  
FT TRANSMEM 581 601  
FT DOMAIN 602 623  
FT TRANSMEM 624 644  
FT DOMAIN 645 653  
FT TRANSMEM 654 674  
FT DOMAIN 675 841  
FT TRANSMEM 842 862  
FT DOMAIN 863 1323  
FT TRANSMEM 278 283  
FT DOMAIN 905 913  
FT TRANSMEM 1030 1035  
FT DOMAIN 1197 1201  
FT SITE 639 639  
FT CARBOHYD 89 89  
FT CARBOHYD 349 349  
FT CARBOHYD 363 363  
FT CARBOHYD 464 464  
FT CARBOHYD 566 566  
SQ SEQUENCE 1323 AA; 142907 MW; 8AE9878F9DD0921 CRC64;  
  
Query Match 69.8%; Score 37; DB 1; Length 1323;  
Best Local Similarity 66.7%; Pred. No. 50;  
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
  
QY 1 FLINPELIV 9  
DB 378 FLVNPVLV 386  
  
RESULT 6  
NME4\_RAT STANDARD; PRT; 1323 AA.  
AC Q62645; Q63381; Q63382; Q63729; Q63730;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Glutamate [NMDA] receptor subunit epsilon 4 precursor (N-methyl  
DE D-aspartate receptor subtype 2D) (NR2D) (NMDAR2D).  
GN GRIN2D.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
OX NCBI\_TaxID=10116;  
RN [1]  
RP SEQUENCE FROM N.A. (ISOFORM 1), AND SEQUENCE OF 1265-1356 FROM N.A.  
RP (ISOFORM 2).  
RC STRAIN=Sprague-Dawley; TISSUE=Forebrain;  
RX MEDLINE=93155102; PubMed=8428958;  
RX Ishii T., Moriyoshi K., Sugihara H., Sakurada K., Kadotani H.,  
RX Yokoi M., Nakazawa C., Shigemoto R., Mizuno N., Nakanishi S.;



"Molecular characterization of the family of the N-methyl-D-aspartate receptor subunits";  
 J. Biol. Chem. 268:2836-2843 (1993).  
 [2]  
 RN SEQUENCE FROM N.A. (ISOFORM 2).  
 RC TISSUE=Brain;  
 RX MEDLINE=94206533; PubMed=7512349;  
 RA Monyer H., Burnashev N., Laurie D.J., Sakmann B., Seeburg P.H.;  
 "Developmental and regional expression in the rat brain and functional properties of four NMDA receptors";  
 Neuron 12:529-540 (1994).  
 [3]  
 RN SEQUENCE FROM N.A. (ISOFORM 2).  
 RC STRAIN=Sprague-Dawley; TISSUE=Brain;  
 RA Boulter J., Pecht G.;  
 Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.  
 CC -!- FUNCTION: NMDA receptor subtype of glutamate-gated ion channels with high calcium permeability and voltage-dependent sensitivity to magnesium. Mediated by glycine.  
 CC -!- SUBUNIT: Heterodimer of an epsilon subunit and a zeta subunit.  
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein.  
 CC -!- ALTERNATIVE PRODUCTS:  
 CC Event=Alternative splicing; Named isoforms=2;  
 CC Name=2;  
 CC IsoId=Q62645-1; Sequence=Displayed;  
 CC Name=1;  
 CC IsoId=Q62645-2; Sequence=VSP 000136;  
 CC Note=No experimental confirmation available;  
 CC -!- TISSUE SPECIFICITY: Expressed in brain, mainly in the subcortical region.  
 CC -!- DEVELOPMENTAL STAGE: Already detected in embryonic stages, peaks at postnatal day 7, and decreases thereafter to adult levels.  
 CC -!- SIMILARITY: Belongs to the ligand-gated ionic channel family.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; D13213; BAA02500.1; -;  
 DR EMBL; D13214; BAA02501.1; -;  
 DR EMBL; L31611; AAC37646.1; -;  
 DR EMBL; L31612; AAC37647.1; -;  
 DR EMBL; U08260; AAA17833.1; -;  
 DR PIR; I78557; I78557.  
 DR HSSP; P19491; 1GR2.  
 DR InterPro; IPR001828; ANF receptor.  
 DR InterPro; IPR001320; Ion\_glu\_receptor.  
 DR InterPro; IPR001508; NMDA\_receptor.  
 DR InterPro; IPR001311; SBP/Glu\_receptor.  
 DR Pfam; PF01094; ANF\_receptor; 1.  
 DR Pfam; PF00060; lig\_chan; 1.  
 DR PRINTS; PR00177; NMDARECEPTOR.  
 DR SMART; SM00079; PBPe; 1.  
 DR Receptor; Signal; Transmembrane; Postsynaptic membrane; Calcium; Glycoprotein; Ionic channel; Magnesium; Alternative splicing.  
 KW Glycoprotein; Ionic channel; Magnesium; Alternative splicing.  
 FT SIGNAL 1 27  
 FT CHAIN 28 1323  
 FT CHAIN 28 1323  
 FT DOMAIN 28 580  
 FT TRANSMEM 581 601  
 FT DOMAIN 602 623  
 FT TRANSMEM 624 644  
 FT DOMAIN 645 653  
 FT TRANSMEM 654 674  
 FT DOMAIN 675 841  
 FT TRANSMEM 842 862  
 FT DOMAIN 863 1323  
 FT DOMAIN 278 283  
 FT DOMAIN 905 913  
 FT POLY-PRO.

FT DOMAIN 1030 1035 POLY-ALA.  
 FT DOMAIN 1197 1201 POLY-PRO.  
 FT SITE 639 639 FUNCTIONAL DETERMINANT OF NMDA RECEPTORS (BY SIMILARITY).  
 FT CARBOHYD 89 89 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 349 349 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 363 363 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 464 464 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 566 566 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT VARSPLIC 1265 1323 CPAAAPTRLTGSRHARRCPHAHMCPLPTASHRRHGG  
 DLGTRGSAHFSSSEV -> RCPCHRTGDTGAGTWAHA  
 GALRISPAWSPRYDAAPPTTAAAPSVASGHGRGAKWT  
 GPSVWGKDRNGPGRTPPEGAASCAPTTFALGEL (in isoform 1).  
 FT /FTId=VSP 000136.  
 FT A -> V (IN REF. 3).  
 FT CONFLICT 25 25  
 FT CONFLICT 47 47 P -> Q (IN REF. 3).  
 FT CONFLICT 67 67 G -> V (IN REF. 2; AAC37646).  
 FT CONFLICT 94 94 R -> P (IN REF. 2).  
 FT CONFLICT 305 305 R -> A (IN REF. 2).  
 FT CONFLICT 635 635 A -> G (IN REF. 3).  
 FT CONFLICT 974 974 E -> D (IN REF. 3).  
 FT CONFLICT 1253 1253 A -> G (IN REF. 3).  
 FT CONFLICT 1266 1267 PR -> TT (IN REF. 3).  
 SQ SEQUENCE 1323 AA; 143100 MW; 40F7D60192579564 CRC64;  
 Query Match 69.8%; Score 37; DB 1; Length 1323;  
 Best Local Similarity 66.7%; Pred. No. 50;  
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
 Qy 1 FLINPELIV 9  
 Db 378 FLVNPISLV 386  
 ID NME4 HUMAN STANDARD; PRT; 1336 AA.  
 AC O15399;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Glutamate [NMDA] receptor subunit epsilon 4 precursor (N-methyl D-aspartate receptor subtype 2D) (NR2D) (NMDAR2D) (EB11).  
 GN GRIN2D.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 [1]  
 RN SEQUENCE FROM N.A.  
 RC TISSUE=Fetal brain;  
 RX MEDLINE=98149394; PubMed=9489750;  
 RA Hess S.D., Daggett L.P., Deal C., Lu C.-C., Johnson E.C., Velicelebi G.;  
 RT "Functional characterization of human N-methyl-D-aspartate subtype 1A/2D receptors";  
 RT J. Neurochem. 70:1269-1279 (1998).  
 CC -!- FUNCTION: NMDA receptor subtype of glutamate-gated ion channels with high calcium permeability and voltage-dependent sensitivity to magnesium. Mediated by glycine.  
 CC -!- SUBUNIT: Heterodimer of an epsilon subunit and a zeta subunit.  
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein.  
 CC -!- SIMILARITY: Belongs to the ligand-gated ionic channel family.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

```

DR EMBL; U77783; AAC15910.1; -.
DR HSSP; P19491; IGR2.
DR GenBank; HGNC:4588; GRIN2D.
DR MIM; 602717; -.
DR InterPro; IPR001828; ANF_receptor.
DR InterPro; IPR001320; Ion_glu_receptor.
DR InterPro; IPR001508; NMDA_receptor.
DR InterPro; IPR001311; SBP/Glu_receptor.
DR Pfam; PF01094; ANF_receptor; 1.
DR Pfam; PF00060; lig_chan; 1.
DR PRINTS; PR00177; NMDARECEPTOR.
DR SMART; SM00079; PBPE; 1.
KW Receptor; Signal; Transmembrane; Postsynaptic membrane; Calcium;
KW Glycoprotein; Ionic channel; Magnesium.
FT SIGNAL 1 27
FT CHAIN 28 1336
FT DOMAIN 28 583
FT TRANSMEM 584 604
FT DOMAIN 605 626
FT TRANSMEM 627 647
FT DOMAIN 648 656
FT TRANSMEM 657 677
FT DOMAIN 678 844
FT TRANSMEM 845 865
FT DOMAIN 866 1336
FT DOMAIN 281 286
FT DOMAIN 908 916
FT DOMAIN 1035 1040
FT DOMAIN 1209 1213
FT DOMAIN 1244 1247
FT SITE 642 642
FT CARBOHYD 92 92
FT CARBOHYD 352 352
FT CARBOHYD 366 366
FT CARBOHYD 467 467
FT CARBOHYD 569 569
FT SEQUENCE 1336 AA; 143558 MW; DECC545F3E416680 CRC64;

Query Match 69.8%; Score 37; DB 1; Length 1336;
Best Local Similarity 66.7%; Pred. No. 51;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 FLINPELIV 9
Db 381 FLVNPISLV 389
|||||
|:|:|:|:|

RESULT 8
DEF2_NITEU STANDARD; PRT; 185 AA.
AC Q82TC8;
DT 15-MAR-2004 (Rel. 43, Created)
DT 15-MAR-2004 (Rel. 43, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Peptide deformylase 2 (EC 3.5.1.88) (PDF 2) (polypeptide deformylase
2).
GN DEF2 OR NE1970.
OS Nitrosomonas europaea.
OC Bacteria; Proteobacteria; Betaproteobacteria; Nitrosomonadales;
OC Nitrosomonadaceae; Nitrosomonas.
OX NCBI_TaxID=915;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 19718 / IPO 14298;
RX MEDLINE=22586410; PubMed=12700255;
RA Chain P., Lamerdin J.E., Larimer F.W., Regala W., Lao V., Land M.,
RA Hauser L., Hooper A.B., Klotz M.G., Norton J., Sayavedra-Soto L.A.,
RA Arclero D.M., Homes N.G., Whittaker M.M., Arp D.J.;
RT "Complete genome sequence of the ammonia-oxidizing bacterium and
RT obligate chemolithoautotroph Nitrosomonas europaea.";
RL J. Bacteriol. 185:2759-2773(2003).

EMBL; BX321862; CAD85881.1; -.
HAMAP; MF_00163; -.
InterPro; IPR000181; Fmet_deformylase.
Pfam; PF01327; Pep_deformylase; 1.
PRINTS; PR01576; PDEFORMYLASE.
ProDom; PD003844; Pep_deformylase; 1.
TIGRFAMs; TIGR00079; pep_deformyl; 1.
KW Protein biosynthesis; Hydrolase; Iron; Complete proteome.
FT ACT_SITE 151 151
FT METAL 108 108
FT METAL 150 150
FT METAL 154 154
FT SEQUENCE 185 AA; 20573 MW; CBELA641426FDEC1 CRC64;

Query Match 67.9%; Score 36; DB 1; Length 185;
Best Local Similarity 77.8%; Pred. No. 9.3;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 LINPELIVS 10
Db 90 LINPELIAS 98
|||||
|:|:|:|:|

RESULT 9
FLIG_BUCBP STANDARD; PRT; 320 AA.
AC Q89AZ9;
DT 15-MAR-2004 (Rel. 43, Created)
DT 15-MAR-2004 (Rel. 43, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Flagellar motor switch protein Flig.
GN FLIG OR BBP069.
OS Buchnera aphidicola (subsp. Baizongia pistaciae).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Buchnera.
OX NCBI_TaxID=135842;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22426901; PubMed=12522265;
RA Van Ham R.C.H.J.; Kamerbeek J.; Palacios C.; Rausell C.; Abascal F.,
RA Bastolla U.; Fernandez J.M.; Jimenez L.; Postigo M.; Silva F.J.;
RA Tamames J.; Viguera E.; Latorre A.; Valencia A.; Moran F.; Moya A.;
RT "Reductive genome evolution in Buchnera aphidicola.";
Proc. Natl. Acad. Sci. U.S.A. 100:581-586(2003).
CC -!- FUNCTION: FLIG IS ONE OF THREE PROTEINS (FLIG, FLIN, FLIM) THAT
CC FORM A SWITCH COMPLEX THAT IS PROPOSED TO BE LOCATED AT THE BASE
CC OF THE BASAL BODY. THIS COMPLEX INTERACTS WITH THE CHEY AND CHEZ
CC CHEMOTAXIS PROTEINS, IN ADDITION TO CONTACTING COMPONENTS OF THE
CC MOTOR THAT DETERMINE THE DIRECTION OF FLAGELLAR ROTATION (BY
CC SIMILARITY).
CC -!- SUBCELLULAR LOCATION: Membrane-associated (By similarity).
CC -!- SIMILARITY: Belongs to the flig family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

```

CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; AB014016; AAC26805.1; -;  
 DR InterPro; IPR000090; Flg\_Motor\_Flag.  
 DR Pfam; PF01706; Flg-C; 1;  
 DR PRINTS; PR00954; FLGMOTOREFLIG  
 KW Chemotaxis; Flagellum; Flagellar rotation; Membrane;  
 KW Complete proteome.  
 SQ SEQUENCE 320 AA; 37070 MW; 82F74CDEB72DE287 CRC64;

Query Match 67.9%; Score 36; DB 1; Length 320;  
 Best Local Similarity 88.9%; Pred. No. 17;  
 Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 LINPELVS 10  
 |||||  
 Db 113 LINPEKVS 121

RESULT 10  
 ID ZW10 DROGR STANDARD; PRT; 714 AA.  
 AC 044219;

DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Centromere/kinetochore protein zw10 (Mitotic 15 protein).  
 GN MIT(1)15 OR ZW10.

OS Drosophila grimshawi (Fruit fly) (Idiomyla grimshawi).  
 CC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 CC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 CC Ephydroidea; Drosophilidae; Drosophila.

OX NCBI\_TaxID=7222;  
 RN [1]  
 RP SEQUENCE FROM N.A.

RX MEDLINE=97444363; PubMed=9298984;  
 RA Starr D.A., Williams B.C., Li Z., Etemad-Moghadam B., Dawe R.K.,  
 RA Goldberg M.L.;  
 RA "Conservation of the centromere/kinetochore protein ZW10.";  
 RL J. Cell Biol. 138:1289-1301(1997).

CC -!- FUNCTION: Required for accurate chromosome segregation (By  
 similarity).

CC -!- SUBCELLULAR LOCATION: EXCLUDED FROM THE NUCLEUS DURING INTERPHASE  
 CC BUT MIGRATES INTO THE NUCLEAR ZONE DURING PROMETAPHASE. AT  
 CC METAPHASE, FOUND IN A FILAMENTOUS STRUCTURE THAT MAY BE  
 CC SPECIFICALLY ASSOCIATED WITH KINETOCHORE MICROTUBULES. AT  
 CC ANAPHASE, FOUND AT OR NEAR KINETOCHORES OF SEPARATING CHROMOSOMES.  
 CC AT THE BEGINNING OF TELOPHASE, BECOMES EXCLUDED AGAIN FROM THE  
 CC NUCLEUS AND IS DISPERSED IN THE CYTOPLASM (BY SIMILARITY).

CC -!- SIMILARITY: Belongs to the ZW10 family.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; U54998; AAB88238.1; -;  
 DR FlyBase; FBgn022892; Dmri\mit(1)15.  
 KW Cell cycle; Meiosis; Mitosis; Nuclear protein; Centromere.  
 SQ SEQUENCE 714 AA; 81262 MW; 62E14636B9D5B12D CRC64;

Query Match 67.9%; Score 36; DB 1; Length 714;  
 Best Local Similarity 100.0%; Pred. No. 41;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 FLINPEL 7

|||||  
 Db 370 FLINPEL 376

RESULT 11

DEF1\_VIBCH STANDARD; PRT; 169 AA.

AC 09KVU3;

DT 28-FEB-2003 (Rel. 41, Created)

DT 28-FEB-2003 (Rel. 41, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Peptide deformylase 1 (EC 3.5.1.88) (PDF 1) (Polypeptide deformylase

1).

GN DEFI OR VC0046.

OS Vibrio cholerae.

OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;

OC Vibrionaceae; Vibrio.

OX NCBI\_TaxID=666;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=E1 Tor N16961 / Serotype O1;

RX MEDLINE=20406833; PubMed=10952301;

RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwinn M.L.,

RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,

RA Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,

RA McDonald L., Utterback T., Fleischmann R.D., Nierman W.C., White O.,

RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,

RA Fraser C.M.;

RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio

cholerae.";

RL Nature 406:477-483(2000).

CC -!- FUNCTION: Removes the formyl group from the N-terminal Met of

CC newly synthesized proteins. Requires at least a dipeptide for an

CC efficient rate of reaction. N-terminal L-methionine is a

CC prerequisite for activity but the enzyme has broad specificity at

CC other positions (By similarity).

CC -!- CATALYTIC ACTIVITY: Formyl-L-methionyl peptide + H(2)O = formate +

CC methionyl peptide.

CC -!- COFACTOR: Binds 1 iron(II) ion (By similarity).

CC -!- SIMILARITY: Belongs to the polypeptide deformylase family.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; AE004095; AAF93224.1; -;

DR PIR; A82373; A82373.

DR HSP; P27251; IDEF.

DR TIGR; VC0046; -;

DR HAMAP; MF\_00163; -; 1.

DR InterPro; IPR000181; Pep\_deformylase.

DR Pfam; PF01327; Pep\_deformylase; 1.

DR PRINTS; PR01576; PDEFORMLASE.

DR ProDom; PD003844; Pep\_deformylase; 1.

DR TIGRFAMS; TIGR00079; pep\_deformyl; 1.

KW Protein biosynthesis; Hydrolase; Iron; Complete proteome.

FT METAL 91 91 IRON (BY SIMILARITY).

FT METAL 133 133 IRON (BY SIMILARITY).

FT ACT SITE 134 134 BY SIMILARITY.

FT METAL 137 137 IRON (BY SIMILARITY).

SQ SEQUENCE 169 AA; 19147 MW; 16DB00B08CA40FC7 CRC64;

Query Match 66.0%; Score 35; DB 1; Length 169;

Best Local Similarity 70.0%; Pred. No. 13;

Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 2 LINPELIVSR 11

|||||

Db 73 LINPEIIKRR 82

RESULT 12

DEF1\_RICCN STANDARD; PRT; 175 AA.

AC Q921Z1;

DT 28-FEB-2003 (Rel. 41, Created)

DT 28-FEB-2003 (Rel. 41, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Peptide deformylase 1 (EC 3.5.1.88) (PDF 1) (Polypeptide deformylase 1).

DE DEF1 OR RC0278.

GN Rickettsia conorii.

OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;

OC Rickettsiaceae; Rickettsiaseae; Rickettsia.

OX NCBI\_TaxID=781;

[1]

RN SEQUENCE FROM N.A.

RP STRAIN=Malish 7;

RC MEDLINE=21442074; PubMed=11557893;

RA Ogata H., Audic S., Renesto-Audiffren P., Fournier P.-E., Barbe V., Samson D., Roux V., Cossart P., Weissenbach J., Claverie J.-M., Raoult D.;

RT "Mechanisms of evolution in Rickettsia conorii and R. prowazekii.;"

RL Science 293:2093-2098(2001).

CC -!- FUNCTION: Removes the formyl group from the N-terminal Met of newly synthesized proteins. Requires at least a dipeptide for an efficient rate of reaction. N-terminal L-methionine is a prerequisite for activity but the enzyme has broad specificity at other positions (By similarity).

CC -!- CATALYTIC ACTIVITY: Formyl-L-methionyl peptide + H(2)O = formate + methionyl peptide.

CC -!- COFACTOR: Binds 1 iron(II) ion (By similarity).

CC -!- SIMILARITY: Belongs to the polypeptide deformylase family.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announcement/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; A5008593; RAL02816.1; -.

DR PIR; F97734; F97734.

DR HAMAP; MF\_00163; -; 1.

DR InterPro; IPR000181; Pep deformylase.

DR Pfam; PF01327; Pep deformylase; 1.

DR PRINTS; PR01576; PDEFORMYLASE.

DR ProDom; PD003844; Pep deformylase; 1.

DR TIGRFAMs; TIGR00079; Pep deformyl; 1.

KW Protein biosynthesis; Hydrolase; Iron; Complete proteome.

FT ACT SITE 142 142 BY SIMILARITY.

FT METAL 99 99 IRON (BY SIMILARITY).

FT METAL 141 141 IRON (BY SIMILARITY).

FT METAL 145 145 IRON (BY SIMILARITY).

SQ SEQUENCE 175 AA; 20279 MW; 93B050C6003C5A6D CRC64;

Query Match 66.0%; Score 35; DB 1; Length 175;

Best Local Similarity 62.5%; Pred. No. 14;

Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 FLINPELI 8

DB 79 FIVNPEII 86

RESULT 13

DEF1\_RICPR STANDARD; PRT; 175 AA.

AC Q9ZDV8;

DT 30-MAY-2000 (Rel. 39, Created)

DT 30-MAY-2000 (Rel. 39, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Peptide deformylase (EC 3.5.1.88) (PDF) (Polypeptide deformylase).

GN DEF OR RP208.

OS Rickettsia prowazekii.

OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;

OC Rickettsiaceae; Rickettsiaseae; Rickettsia.

OX NCBI\_TaxID=782;

[1]

RN SEQUENCE FROM N.A.

RP STRAIN=Madrid E;

RC MEDLINE=99039499; PubMed=9823893;

RA Andersson S.G.E., Zomorodipour A., Andersson J.O., Sichertitz-Ponten T., Alsmark U.C.M., Podowski R.M., Naesslund A.K., Eriksson A.-S., Winkler H.H., Kurland C.G.;

RT "The genome sequence of Rickettsia prowazekii and the origin of mitochondria.;"

RL Nature 396:133-140(1998).

CC -!- FUNCTION: Removes the formyl group from the N-terminal Met of newly synthesized proteins. Requires at least a dipeptide for an efficient rate of reaction. N-terminal L-methionine is a prerequisite for activity but the enzyme has broad specificity at other positions (By similarity).

CC -!- CATALYTIC ACTIVITY: Formyl-L-methionyl peptide + H(2)O = formate + methionyl peptide.

CC -!- COFACTOR: Binds 1 iron(II) ion (By similarity).

CC -!- SIMILARITY: Belongs to the polypeptide deformylase family.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announcement/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; AJ235270; CAA14673.1; -.

DR PIR; B71732; B71732.

DR HSP; P27251; 1DEF.

DR HAMAP; MF\_00163; -; 1.

DR InterPro; IPR000181; Pep deformylase.

DR Pfam; PF01327; Pep deformylase; 1.

DR PRINTS; PR01576; PDEFORMYLASE.

DR ProDom; PD003844; Pep deformylase; 1.

DR TIGRFAMs; TIGR00079; Pep deformyl; 1.

KW Protein biosynthesis; Hydrolase; Iron; Complete proteome.

FT ACT SITE 142 142 BY SIMILARITY.

FT METAL 99 99 IRON (BY SIMILARITY).

FT METAL 141 141 IRON (BY SIMILARITY).

FT METAL 145 145 IRON (BY SIMILARITY).

SQ SEQUENCE 175 AA; 20269 MW; 2400553B491ABED9 CRC64;

Query Match 66.0%; Score 35; DB 1; Length 175;

Best Local Similarity 62.5%; Pred. No. 14;

Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 FLINPELI 8

DB 79 FIVNPEII 86

RESULT 14

IF2G\_YEAST STANDARD; PRT; 527 AA.

AC P32481;

DT 01-OCT-1993 (Rel. 27, Created)

DT 01-OCT-1993 (Rel. 27, Last sequence update)

DT 15-MAR-2004 (Rel. 43, Last annotation update)

DE Eukaryotic translation initiation factor 2 gamma subunit (eIF-2-gamma).

DE GCD11 OR TIF213 OR YER025W.

GN Saccharomyces cerevisiae (Baker's yeast).

OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 26, 2004, 18:18:25 ; Search time 0.877872 Seconds  
(without alignments)

3953.539 Million cell updates/sec

Title: US-09-302-812-13

Perfect score: 53

Sequence: 1 FLINPELIVSR 11

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL.25.\*

1: sp\_archaea.\*

2: sp\_bacteria.\*

3: sp\_fungi.\*

4: sp\_human.\*

5: sp\_invertebrate.\*

6: sp\_mammal.\*

7: sp\_mhc.\*

8: sp\_organelle.\*

9: sp\_phase.\*

10: sp\_plant.\*

11: sp\_rodent.\*

12: sp\_virus.\*

13: sp\_vertebrate.\*

14: sp\_unclassified.\*

15: sp\_rvirus.\*

16: sp\_bacteriap.\*

17: sp\_archaeap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	53	100.0	920	11 Q8CB72	Q8cb72 mus musculus
2	53	100.0	961	11 Q80YQ6	Q80yq6 mus musculus
3	53	100.0	968	11 Q88622	Q88622 mus musculus
4	53	100.0	972	11 Q9QYM2	Q9qym2 rattus norv
5	53	100.0	977	6 Q02776	Q02776 bos taurus
6	52	98.1	976	4 Q9Y4W7	Q9y4w7 homo sapien
7	52	98.1	976	4 Q86W56	Q86w56 homo sapien
8	52	98.1	976	4 Q7Z742	Q7z742 homo sapien
9	42	79.2	364	10 Q9SKB4	Q9skb4 arabidopsis
10	41	77.4	249	16 Q9Z7P9	Q9z7p9 chlamydia p
11	39	73.6	312	2 Q8KLF8	Q8klf8 rhizobium e
12	38	71.7	150	16 Q895Q2	Q895q2 clostridium
13	38	71.7	386	16 Q82VH5	Q82vh5 nitrosomona
14	38	71.7	472	16 Q8FA62	Q8faf62 leptospira
15	38	71.7	548	10 Q9SKB3	Q9skb3 arabidopsis
16	38	71.7	1063	5 Q8SVI4	Q8svi4 encephalito

17	37	69.8	161	12 Q9DVM2	Q9dvw2 pluteella xy
18	37	69.8	302	3 Q04110	Q04110 saccharomyc
19	37	69.8	522	10 Q8VYAL	Q8vyal arabidopsis
20	37	69.8	13414	5 Q81616	Q81616 cryptospori
21	36	67.9	185	16 Q82TC8	Q82tc8 nitrosomona
22	36	67.9	194	16 Q7V1K9	Q7v1k9 prochloroco
23	36	67.9	200	16 Q9K7M6	Q9k7m6 bacillus ha
24	36	67.9	219	16 Q9KRK2	Q9krk2 vibrio chol
25	36	67.9	219	16 Q87QD7	Q87qd7 vibrio para
26	36	67.9	230	2 Q9EYK4	Q9eyk4 lactobacill
27	36	67.9	255	16 Q88X77	Q88x77 lactobacill
28	36	67.9	314	16 Q92XV5	Q92xv5 rhizobium m
29	36	67.9	320	16 Q89AZ9	Q89az9 buchnera ap
30	36	67.9	519	10 Q48655	Q48655 oryza sativ
31	36	67.9	520	10 Q84S50	Q84s50 oryza sativ
32	36	67.9	536	10 Q94315	Q94315 oryza sativ
33	36	67.9	557	16 Q9XOV0	Q9xov0 thermotoga
34	36	67.9	688	16 Q8A213	Q8a213 bacteroides
35	35	66.0	160	16 Q9KKN4	Q9kkn4 vibrio chol
36	35	66.0	160	17 Q8TISA	Q8tis4 methanosarc
37	35	66.0	197	16 Q7VBM4	Q7vbm4 prochloroco
38	35	66.0	208	16 Q97M04	Q97m04 clostridium
39	35	66.0	226	17 Q97X84	Q97x84 sulfolobus
40	35	66.0	246	16 Q99T40	Q99t40 staphylococ
41	35	66.0	246	16 Q8NVV9	Q8nvv9 staphylococ
42	35	66.0	297	2 Q53036	Q53036 haemophilus
43	35	66.0	321	17 Q974E9	Q974e9 sulfolobus
44	35	66.0	343	16 Q9RZB5	Q9rzb5 deinococcus
45	35	66.0	377	17 Q9YBM2	Q9ybm2 aeropyrum p

## ALIGNMENTS

RESULT 1

Q8CB72 PRELIMINARY; PRT; 920 AA.  
AC Q8CB72;  
DT 01-MAR-2003 (TREMBLrel. 23, Created)  
DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)  
DT 01-JUN-2003 (TREMBLrel. 24, Last annotation update)  
DE Poly.  
GN PARG.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6J; TISSUE=Bone;  
RX MEDLINE=22354683; PubMed=12466851;  
RA The FANTOM Consortium,  
RA the RIKEN Genome Exploration Research Group Phase I & II Team;  
RT "Analysis of the mouse transcriptome based on functional annotation of  
RT 60,770 full-length cDNAs.";  
RL Nature 420:563-573(2002).  
DR ENBL; AK036656; BAC29519.1; -.  
DR MGD; MGI:1347094; Parg.  
DR InterPro; IPR007724; PARG.  
DR Pfam; PF05028; PARG; 1.  
SQ SEQUENCE 920 AA; 103799 MW; 379E779E121B09A1 CRC64;

Query Match 100.0%; Score 53; DB 11; Length 920;  
Best Local Similarity 100.0%; Pred. No. 0.16; Indels 0; Gaps 0;  
Matches 11; Conservative 0; Mismatches 0;

Qy 1 FLINPELIVSR 11

Db 752 FLINPELIVSR 762

RESULT 2

Q80YQ6

```

ID Q80YQ6 PRELIMINARY; PRT; 961 AA.
AC Q80YQ6;
DT 01-JUN-2003 (TReMBLrel. 24, Created)
DT 01-JUN-2003 (TReMBLrel. 24, Last sequence update)
DT 01-JUN-2003 (TReMBLrel. 25, Last annotation update)
DE Poly(ADP-ribose) glycohydrolase.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OC NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6; TISSUE=Brain;
RA Strausberg R.;
RL Submitted (APR-2003) to the EMBL/GenBank/DBSJ databases.
DR EMBL; BC050892; AAH50892.1; -.
DR GO; GO:0016787; F:hydrolase activity; IEA.
DR InterPro; IPR007724; PARG.
DR Pfam; PF05028; PARG; 1.
KW Hydrolase.
SQ SEQUENCE 961 AA; 108580 MW; 1558F120C367CF63 CRC64;

Query Match 100.0%; Score 53; DB 11; Length 961;
Best Local Similarity 100.0%; Pred. No. 0.17;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FLINPELIVSR 11
Db |||||
752 FLINPELIVSR 762

RESULT 3
ID Q86G22 PRELIMINARY; PRT; 968 AA.
AC Q86G22;
DT 01-NOV-1998 (TReMBLrel. 08, Created)
DT 01-NOV-1998 (TReMBLrel. 08, Last sequence update)
DT 01-JUN-2003 (TReMBLrel. 24, Last annotation update)
DE Poly(ADP-ribose) glycohydrolase.
GN PARG.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OC NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99380098; PubMed=10449915;
RA Ame J.C., Apio F., Jacobson E.L., Jacobson M.K.;
RT "Assignment of the poly(ADP-ribose) glycohydrolase gene (PARG) to
RT human chromosome 10q11.23 and mouse chromosome 14B by in situ
RT hybridization."
RL Cytogenet. Cell Genet. 85:269-270 (1999).
DR EMBL; AF079557; AAC28735.1; -.
DR MGD; MGI:1347094; Parg.
DR GO; GO:0016787; F:hydrolase activity; IEA.
DR InterPro; IPR007724; PARG.
DR Pfam; PF05028; PARG; 1.
KW Hydrolase.
SQ SEQUENCE 968 AA; 109169 MW; 3981DA4809E1E007 CRC64;

Query Match 100.0%; Score 53; DB 11; Length 968;
Best Local Similarity 100.0%; Pred. No. 0.17;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FLINPELIVSR 11
Db |||||
751 FLINPELIVSR 761

RESULT 4
ID Q9QYM2 PRELIMINARY; PRT; 972 AA.
AC Q9QYM2;

```

```

DT 01-MAY-2000 (TReMBLrel. 13, Created)
DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)
DT 01-JUN-2003 (TReMBLrel. 24, Last annotation update)
DE Poly(ADP-ribose) glycohydrolase.
GN PARG.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OC NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BUF; TISSUE=Colon;
RA Shimokawa T., Masutani M., Nagasawa S., Nozaki T., Ikota N., Araki S.,
RA Aoki Y., Nakagawa H., Sugimura T.;
RT "Isolation and cloning of rat poly(ADP-ribose) glycohydrolase
RT (Parg).";
RL Submitted (OCT-1998) to the EMBL/GenBank/DBSJ databases.
DR EMBL; AB019366; BAAS7901.1; -.
DR GO; GO:0016787; F:hydrolase activity; IEA.
DR InterPro; IPR007724; PARG.
DR Pfam; PF05028; PARG; 1.
KW Hydrolase.
SQ SEQUENCE 972 AA; 109002 MW; 3DED7885684F59E9 CRC64;

Query Match 100.0%; Score 53; DB 11; Length 972;
Best Local Similarity 100.0%; Pred. No. 0.17;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FLINPELIVSR 11
Db |||||
755 FLINPELIVSR 765

RESULT 5
ID Q02776 PRELIMINARY; PRT; 977 AA.
AC Q02776;
DT 01-JUL-1997 (TReMBLrel. 04, Created)
DT 01-JUL-1997 (TReMBLrel. 04, Last sequence update)
DT 01-JUN-2003 (TReMBLrel. 24, Last annotation update)
DE Poly(ADP-ribose) glycohydrolase.
GN BPARG.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OC NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97277328; PubMed=9115250;
RA Lin W., Ame J.C., Aboul-Ela N., Jacobson E.L., Jacobson M.K.;
RT "Isolation and characterization of the cDNA encoding bovine poly(ADP-
RT ribose) glycohydrolase."
RL J. Biol. Chem. 272:11895-11901 (1997).
DR EMBL; U78975; AAB53370.1; -.
DR GO; GO:0016787; F:hydrolase activity; IEA.
DR InterPro; IPR007724; PARG.
DR Pfam; PF05028; PARG; 1.
KW Hydrolase.
SQ SEQUENCE 977 AA; 110837 MW; 87D2100F979DF377 CRC64;

Query Match 100.0%; Score 53; DB 6; Length 977;
Best Local Similarity 100.0%; Pred. No. 0.17;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FLINPELIVSR 11
Db |||||
760 FLINPELIVSR 770

RESULT 6
ID Q9Y4W7 PRELIMINARY; PRT; 976 AA.

```

AC Q9Y4W7;  
 DT 01-NOV-1999 (TrEMBLrel. 12, Created)  
 DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
 DE Poly(ADP-ribose) glycohydrolase.  
 GN HPARG.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RN SEQUENCE FROM N.A.  
 RX MEDLINE=99380098; PubMed=10449915;  
 RA Ame J.C., Apion F., Jacobson E.L., Jacobson M.K.;  
 RT "Assignment of the poly(ADP-ribose) glycohydrolase gene (PARG) to  
 RT human chromosome 10q11.23 and mouse chromosome 14B by in situ  
 RT hybridization.";  
 RL Cytogenet. Cell Genet. 85:269-270(1999).  
 DR EMBL; AF005043; AAB61614.1; -;  
 DR Genbank; HGNC:8605; PARG.  
 DR GO; GO:0005737; C:cytoplasm; TAS.  
 DR GO; GO:0004649; F:poly(ADP-ribose) glycohydrolase activity; TAS.  
 DR InterPro; IPR007724; PARG.  
 DR Pfam; PF05028; PARG; 1.  
 KW Hydrolase.  
 SQ SEQUENCE 976 AA; 110903 MW; 42A0F51466FCE7DC CRC64;  
 Query Match 98.1%; Score 52; DB 4; Length 976;  
 Best Local Similarity 90.9%; Pred. No. 0.27;  
 Matches 10; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 FLINPELIVSR 11  
 DB 759 FLINPELIIISR 769  
 RESULT 7  
 Q86W56 PRELIMINARY; PRT; 976 AA.  
 AC Q86W56;  
 DT 01-JUN-2003 (TrEMBLrel. 24, Created)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Similar to poly(ADP-ribose) glycohydrolase.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RN SEQUENCE FROM N.A.  
 RC TISSUE=Skin;  
 RA Strausberg R.;  
 RL Submitted (APR-2003) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; BC050560; AAH50560.1; -;  
 DR GO; GO:0016787; F:hydrolase activity; IEA.  
 DR InterPro; IPR007724; PARG.  
 DR Pfam; PF05028; PARG; 1.  
 KW Hydrolase.  
 SQ SEQUENCE 976 AA; 111109 MW; D6646353C6D0180E CRC64;  
 Query Match 98.1%; Score 52; DB 4; Length 976;  
 Best Local Similarity 90.9%; Pred. No. 0.27;  
 Matches 10; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 FLINPELIVSR 11  
 DB 759 FLINPELIIISR 769  
 RESULT 8  
 Q72742 PRELIMINARY; PRT; 976 AA.  
 AC Q72742;

DT 01-OCT-2003 (TrEMBLrel. 25, Created)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Hypothetical protein.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RN SEQUENCE FROM N.A.  
 RC TISSUE=Testis;  
 RX MEDLINE=22388257; PubMed=12477932;  
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stapleton M., Soares M.B., Donald M.F., Casavant T.L., Scheetz T.E.,  
 RA Brownstein M.J., Udwin T.B., Toshiyuki S., Carninci P., Prange C.,  
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,  
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahney J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,  
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,  
 RA Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,  
 RA Jones S.J., Marra M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length human  
 RT and mouse cDNA sequences.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 RN [2]  
 RN SEQUENCE FROM N.A.  
 RC TISSUE=Testis;  
 RA Strausberg R.;  
 RL Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; BC052966; AAH52966.1; -;  
 KW Hypothetical protein.  
 SQ SEQUENCE 976 AA; 111090 MW; 47C8EE1826C4A74A CRC64;  
 Query Match 98.1%; Score 52; DB 4; Length 976;  
 Best Local Similarity 90.9%; Pred. No. 0.27;  
 Matches 10; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 FLINPELIVSR 11  
 DB 759 FLINPELIIISR 769  
 RESULT 9  
 Q9SKB4 PRELIMINARY; PRT; 364 AA.  
 AC Q9SKB4;  
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
 DE Putative poly(ADP-ribose) glycohydrolase.  
 GN AT2G31860.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.  
 OX NCBI\_TaxID=3702;  
 RN [1]  
 RN SEQUENCE FROM N.A.  
 RC STRAIN=cv. Columbia;  
 RX MEDLINE=20083487; PubMed=10617197;  
 RA Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,  
 RA Fujii C.-Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblum T.V.,  
 RA Buell C.R., Ketchum K.A., Lee J.-J., Ronning C.M., Koo H., Moffat K.S.,  
 RA Cronin L.A., Shen M., VanAken S.E., Unayam L., Tallon L.J., Gill J.E.,  
 RA Adams M.D., Carrera A.J., Creasy T.H., Goodman H.M., Somerville C.R.,



RA Copenhaver G.P., Preuss D., Niernan W.C., White O., Eisen J.A.,  
 RA Salzberg S.L., Fraser C.M., Venter J.C.;  
 RT "Sequence and analysis of chromosome 2 of the plant Arabidopsis  
 RT thaliana.";  
 RL Nature 402:761-768(1999).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. Columbia;  
 RA Lin X.;  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AC006533; AAD32286.1; -;  
 DR PIR; A84726; A84726.  
 DR GO; GO:0016787; F:hydrolyase activity; IEA.  
 DR InterPro; IPR007724; PARG.  
 DR Pfam; PF05028; PARG; 1.  
 KW Hydrolyase.  
 SQ SEQUENCE 364 AA; 40896 MW; 5DECA51A72089A94 CRC64;

Query Match 79.2%; Score 42; DB 10; Length 364;  
 Best Local Similarity 88.9%; Pred. No. 8.8;  
 Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 FLINPELIV 9  
 :|||||  
 Db 202 FMINPELIV 210

## RESULT 10

Q927P9 PRELIMINARY; PRT; 249 AA.  
 AC Q927P9;  
 DT 01-MAY-1999 (TREMBlrel. 10, Created)  
 DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)  
 DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)  
 DE DNA polymerase III epsilon chain.  
 GN DNAQ 2 OR CPN0655 OR CP0092 OR CPB0681.  
 OS Chlamydia pneumoniae (Chlamydia pneumoniae).  
 OC Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydia.  
 OX NCBI\_TaxID=83558;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CWL029;  
 RX MEDLINE=99206606; PubMed=10192388;  
 RA Kalman S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,  
 RA Olinger L., Grimwood J., Davis R.W., Stephens R.S.;  
 RT "Comparative genomes of Chlamydia pneumoniae and C. trachomatis.";  
 RL Nat. Genet. 21:385-389(1999).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=AR39;  
 RX MEDLINE=20150255; PubMed=10684935;  
 RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,  
 RA White O., Hickey E.K., Peterson J., Utterback T., Berry K., Bass S.,  
 RA Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,  
 RA Gwin M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,  
 RA Eisen J., Fraser C.M.;  
 RT "Genome sequences of Chlamydia trachomatis MoPn and Chlamydia  
 RT pneumoniae AR39.";  
 RL Nucleic Acids Res. 28:1397-1406(2000).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=J138;  
 RX MEDLINE=20330349; PubMed=10871362;  
 RA Shirai M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,  
 RA Shiba T., Ishii K., Hattori M., Kuhara S., Nakazawa T.;  
 RT "Comparison of whole genome sequences of Chlamydia pneumoniae J138  
 RT from Japan and CWL029 from USA.";  
 RL Nucleic Acids Res. 28:2311-2314(2000).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=TW-183;  
 RA Geng M.M., Schuhmacher A., Muehlendorfer I., Bensch K.W., Schaefer K.P.,  
 RA Schneider S., Pohl T., Essig A., Marre R., Melchers K.;

RT "The genome sequence of Chlamydia pneumoniae TW183 and comparison with  
 RT other Chlamydia strains based on whole genome sequence analysis.";  
 RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AE001648; AAD18794.1; -;  
 DR EMBL; AE002172; AAF73628.1; -;  
 DR EMBL; AP002547; BAA9862.1; -;  
 DR EMBL; AE017159; AAP98610.1; -;  
 DR PIR; D86572; D86572.  
 DR PIR; H72051; H72051.  
 DR TIGR; CP0092; -;  
 DR GO; GO:0005622; C:intracellular; IEA.  
 DR GO; GO:0003887; F:DNA-directed DNA polymerase activity; IEA.  
 DR GO; GO:0004537; F:exonuclease activity; IEA.  
 DR GO; GO:0006260; P:DNA replication; IEA.  
 DR InterPro; IPR006054; DnaQ.  
 DR InterPro; IPR006055; Exonuclease.  
 DR Pfam; PF00929; Exonuclease; 1.  
 DR SMART; SM00479; EXOIII; 1.  
 DR TIGRFAMs; TIGR00573; dnaq; 1.  
 KW Complete proteome.  
 SQ SEQUENCE 249 AA; 28718 MW; E4B429A4BB895705 CRC64;

Query Match 77.4%; Score 41; DB 16; Length 249;  
 Best Local Similarity 80.0%; Pred. No. 9.3;  
 Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 FLINPELIVS 10  
 :|||||:|  
 Db 45 FLINPERVVS 54

## RESULT 11

Q8KLF8 PRELIMINARY; PRT; 312 AA.  
 AC Q8KLF8;  
 DT 01-OCT-2002 (TREMBlrel. 22, Created)  
 DT 01-OCT-2002 (TREMBlrel. 22, Last sequence update)  
 DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)  
 DE Hypothetical protein.  
 GN NODD1.  
 OS Rhizobium etli.  
 OG Plasmid symbiotic plasmid p42D.  
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;  
 OC Rhizobiaceae; Rhizobium/Agrobacterium group; Rhizobium.  
 OX NCBI\_TaxID=29449;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CFN42;  
 RX MEDLINE=91193195; PubMed=2013564;  
 RA Girard M.L., Flores M., Brom S., Romero D., Palacios R., Davila G.;  
 RT "Structural complexity of the symbiotic plasmid of Rhizobium  
 RT leguminosarum bv. phaseoli.";  
 RL J. Bacteriol. 173:2411-2419(1991).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CFN42;  
 RX MEDLINE=97419521; PubMed=9274036;  
 RA Ramirez-Romero M.A., Bustos P., Girard L., Rodriguez O.,  
 RA Cevallos M.A., Davila G.;  
 RT "Sequence, localization and characteristics of the replicator region  
 RT of the symbiotic plasmid of Rhizobium etli.";  
 RL Microbiology 143:2825-2831(1997).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CFN42;  
 RA Quintero V., Cevallos M.A., Davila G.;  
 RT "A site-specific recombinase and RecA are required to exert  
 RT incompatibility towards the symbiotic plasmid of Rhizobium etli.";  
 RL Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.  
 CC -1- SIMILARITY: BELONGS TO THE LYSR FAMILY OF TRANSCRIPTIONAL  
 CC REGULATORS.  
 DR EMBL; U80928;  
 DR GO; GO:0046821; C:extrachromosomal DNA; IEA.

DR GO; GO:0003700; F:transcription factor activity; IEA.  
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.  
DR GO; GO:0006350; P:transcription; IEA.  
DR InterPro; IPR000847; HTH\_LysR.  
DR Pfam; PF00126; HTH 1; 1.  
DR PRINTS; PR00039; HTHLYSR.  
DR PROSITE; PS00044; HTH\_LYSR FAMILY; 1.  
DR Hypothetical protein; DNA-Binding; Transcription;  
KW Transcription regulation; Plasmid.  
SQ SEQUENCE 312 AA; 35248 MW; B81BF4CA7B5396DC CRC64;

Query Match 73.6%; Score 39; DB 2; Length 312;  
Best Local Similarity 72.7%; Pred. No. 29;  
Matches 8; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 FLINPELIVSR 11  
||| |||||:  
Db 148 FLIPELLMSR 158

## RESULT 12

Q895Q2 PRELIMINARY; PRT; 150 AA.  
AC Q895Q2;  
DT 01-JUN-2003 (TrEMBLrel. 24, Created)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Polypeptide deformylase (EC 3.5.1.31).  
GN CTC01219.  
OS Clostridium tetani.  
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;  
OC Clostridium.  
OX NCBI\_TaxID=1513;  
RN [1]

RP SEQUENCE FROM N.A.  
RX STRAIN=Massachusetts / E88;  
RX MEDLINE=22457253; PubMed=12552129;  
RA Bruggemann H., Bauner S., Fricke W.F., Wierzer A., Liesegang H.,  
RA Decker I., Herzberg C., Martinez-Arias R., Merkl R., Henne A.,  
RA Gottschalk G.;  
RT "The genome sequence of Clostridium tetani, the causative agent of tetanus disease."  
RL Proc. Natl. Acad. Sci. U.S.A. 100:1316-1321 (2003).

DR EMBL; AE015940; AAC3788.1; ...  
DR GO; GO:0008463; F:formylmethionine deformylase activity; IEA.  
DR GO; GO:0016787; F:hydrolase activity; IEA.  
DR GO; GO:0006412; P:protein biosynthesis; IEA.  
DR InterPro; IPR000181; Pep deformylase.  
DR Pfam; PF01327; Pep deformylase; 1.  
DR PRINTS; PR01576; PDEFORMYLASE.  
DR Hydrolase; Complete proteome.  
SQ SEQUENCE 150 AA; 16924 MW; 2C0492B368DC8E7 CRC64;

Query Match 71.7%; Score 38; DB 16; Length 150;  
Best Local Similarity 87.5%; Pred. No. 21;  
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 FLINPELI 8  
||| |||||:  
Db 69 FLINPEII 76

## RESULT 13

Q82VHS PRELIMINARY; PRT; 386 AA.  
AC Q82VHS;  
DT 01-JUN-2003 (TrEMBLrel. 24, Created)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Hyd family secretion protein.  
GN NE1113.  
OS Nitrosomonas europaea.  
OC Bacteria; Proteobacteria; Betaproteobacteria; Nitrosomonadales;

OC Nitrosomonadaceae; Nitrosomonas.  
OX NCBI\_TaxID=915;  
RN [1]

RP SEQUENCE FROM N.A.  
RX STRAIN=ATCC 19718 / IFO 14298;  
RX MEDLINE=22586410; PubMed=12700255;  
RA Chain P., Lamerdin J.E., Larimer F.W., Regala W., Lao V., Land M.,  
RA Hauser L., Hooper A.B., Klotz M.G., Norton J., Sayavedra-Soto L.A.,  
RA Arciero D.M., Hommes N.G., Whittaker M.M., Arp D.J.;  
RT "Complete genome sequence of the ammonia-oxidizing bacterium and  
RT obligate chemolithoautotroph Nitrosomonas europaea.";  
RL J. Bacteriol. 185:2759-2773 (2003).

DR EMBL; BX321859; CAD85024.1; ...  
DR GO; GO:0016020; C:membrane; IEA.  
DR GO; GO:0008565; F:protein transporter activity; IEA.  
DR GO; GO:0009306; P:protein secretion; IEA.  
DR InterPro; IPR006143; HlyD.  
DR InterPro; IPR000437; Prok\_lipoprot\_S.  
DR Pfam; PF0529; HlyD; 1.  
DR PROSITE; PS00013; PROKAR\_LIPOPROTEIN; 1.  
KW Complete proteome.  
SQ SEQUENCE 386 AA; 42412 MW; 626720F804964CFE CRC64;

Query Match 71.7%; Score 38; DB 16; Length 386;  
Best Local Similarity 87.5%; Pred. No. 57;  
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 FLINPELI 8  
||| |||||:  
Db 93 FLINPELL 100

## RESULT 14

Q8F462 PRELIMINARY; PRT; 472 AA.  
AC Q8F462;  
DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Tld protein.  
GN LA2185.  
OS Leptospira interrogans.  
OC Bacteria; Spirochaetes; Spirochaetales; Leptospiraceae; Leptospira.  
OX NCBI\_TaxID=173;  
RN [1]

RP SEQUENCE FROM N.A.  
RX STRAIN=56601 / Serogroup Icterohaemorrhagiae / Serovar lai;  
RA Ren S.;  
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AE011389; AAN49384.1; ...  
DR InterPro; IPR002510; Peptidase\_U62.  
DR Pfam; PF01523; PMDA\_TldB; 1.  
KW Complete proteome.  
SQ SEQUENCE 472 AA; 51527 MW; 6D72FE8434FD59B6 CRC64;

Query Match 71.7%; Score 38; DB 16; Length 472;  
Best Local Similarity 100.0%; Pred. No. 70;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 INPELIVS 10  
||| |||||:  
Db 229 INPELIVS 236

## RESULT 15

Q9SKB3 PRELIMINARY; PRT; 548 AA.  
ID Q9SKB3  
AC Q9SKB3; Q94ET7;  
DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Putative poly(ADP-ribose) glycohydrolase.  
GN AT2G31870 OR TEJ.

OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.  
 OC NCBI\_taxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. Columbia;  
 RA Lin X., Kaul S., Shea T.P., Fujii C.Y., Shen M., VanAken S.E.,  
 RA Barnstead M.E., Mason T.M., Bowman C.L., Ronning C.M., Benito M.-I.,  
 RA Carrera A.J., Creasy T.H., Buell C.R., Town C.D., Nierman W.C.,  
 RA Fraser C.M., Venter J.C.;  
 RL Submitted (FEB-2002) to the EMBL/GenBank/DDbj databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Panda S., Poirier G.G., Kay S.A.;  
 RT "TEJ defines a role for poly-ADP-ribosylation in establishing period  
 RT length of the Arabidopsis circadian oscillator";  
 RL Submitted (JUN-2001) to the EMBL/GenBank/DDbj databases.  
 DR EMBL; AC006533; AAD32285.2; -;  
 DR EMBL; AF394690; AAK72256.1; -;  
 DR PIR; B84726; B84726.  
 DR GO; GO:0016787; F:hydrolase activity; IEA.  
 DR InterPro; IPR007724; PARG.  
 DR Pfam; PF05028; PARG; 1.  
 KW Hydrolase.  
 SQ SEQUENCE 548 AA; 62169 MW; F1A79FDA157C3329 CRC64;  
 Query Match 71.7%; Score 38; DB 10; Length 548;  
 Best Local Similarity 87.5%; Pred. No. 82;  
 Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 FLINPELI 8  
 Db 277 FWINPELI 284

Search completed: May 26, 2004, 18:46:12  
 Job time : 2.87787 secs

GenCore version 5.1.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 26, 2004, 17:19:25 ; Search time 1.83313 Seconds  
(without alignments)  
2466.140 Million cell updates/sec

Title: US-09-302-812-14  
Perfect score: 71  
Sequence: 1 IALXLPNIXTOPIPLL 16

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_29Jan04:\*  
1: Geneseqp1980s:\*  
2: Geneseqp1990s:\*  
3: Geneseqp2000s:\*  
4: Geneseqp2001s:\*  
5: Geneseqp2002s:\*  
6: Geneseqp2003as:\*  
7: Geneseqp2003bs:\*  
8: Geneseqp2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	67	94.4	16	5	Aae25637 Bovine po
2	67	94.4	16	5	Aau76028 Bovine po
3	67	94.4	16	5	Aau76019 Bovine po
4	67	94.4	16	6	Abg72286 Oligopept
5	65	91.5	17	5	Aae25649 Bovine po
6	65	91.5	968	5	Aae25631 Murine po
7	65	91.5	968	5	Aau76022 Mouse po
8	65	91.5	968	5	Aau76013 Mouse pol
9	65	91.5	968	6	Abg72280 Murine p
10	65	91.5	976	5	Aae25630 Human pol
11	65	91.5	976	5	Aau76021 Human pol
12	65	91.5	976	5	Aau76012 Human pol
13	65	91.5	976	6	Abg72279 Human pol
14	65	91.5	977	5	Aae25629 Bovine po
15	65	91.5	977	5	Aau76020 Bovine po
16	65	91.5	977	5	Aau75799 Bovine po
17	65	91.5	977	6	Abg72278 Bovine po
18	46	64.8	768	4	Abbs59491 Drosophil
19	46	64.8	768	5	Aae25632 Fruit fly
20	46	64.8	768	5	Aau76023 Fruit fly
21	46	64.8	768	5	Aau76014 Fruit fly
22	46	64.8	768	6	Abg72281 Fruit fly
23	43	60.6	279	3	Aag46848 Arabidops
24	43	60.6	309	3	AAG46847 Arabidops
25	43	60.6	521	6	ADAI5481 A. thalia

26	43	60.6	521	7	ADD5676	Thalacres
27	43	60.6	521	7	ADD30889	Plant yie
28	41	57.7	961	7	ADB47689	A. gossyp
29	40	56.3	429	6	ABU30155	Protein e
30	40	56.3	774	4	ABB71927	Drosophil
31	39	54.9	109	6	ABP76255	Human GEN
32	39	54.9	223	4	ABG04588	Novel hum
33	39	54.9	228	7	ADB82748	Human pro
34	39	54.9	335	4	ABG09631	Novel hum
35	39	54.9	423	5	AAO14067	Human lys
36	39	54.9	423	7	ADE61363	Human Pro
37	39	54.9	423	7	AD861361	Rat Prote
38	39	54.9	821	4	ABG25804	Novel hum
39	38	53.5	56	4	AAH81219	Human hae
40	38	53.5	56	4	AAH80769	Human hae
41	38	53.5	56	4	AAH81836	Human hae
42	38	53.5	102	5	ABP09179	Human ORF
43	38	53.5	220	3	ABE21243	DaEPV par
44	38	53.5	329	5	ABB47825	Listeria
45	38	53.5	329	6	ABU33015	Protein e

ALIGNMENTS

RESULT 1  
AAE25637  
ID AAE25637 standard; peptide; 16 AA.  
XX  
AC AAE25637;  
XX  
DT 04-NOV-2002 (first entry)  
XX  
DE Bovine poly adenosine diphosphate-ribose glycohydrolase oligopeptide #4.  
XX  
KW Bovine; poly adenosine diphosphate-ribose glycohydrolase; bPARG; enzyme;  
KW ADP-ribose; neoplastic disorder; genotoxic oxidative stress; nontropic;  
KW cardiac disorder; neuronal disorder; reperfusion injury; neurotoxicity;  
KW Alzheimer's disease; Huntington's disease; Parkinson's disease; cardiant;  
KW cytostatic; vasotropic; neuroprotective; anticonvulsant; gene therapy;  
KW antisense therapy.  
XX  
OS Bos taurus.  
XX  
FH Key Location/Qualifiers  
FT Misc-difference 4 /label= Unknown  
FT Misc-difference 9 /label= Unknown  
FT  
FT US6395543-B1.  
PN  
PD 28-MAY-2002.  
XX  
PF 23-FEB-2000; 2000US-00511507.  
PR  
PR 01-MAY-1998; 98US-0083768P.  
PR 30-APR-1999; 99US-00302812.  
XX  
XX (KENT ) UNIV KENTUCKY RES FOUND.  
PA  
XX Jacobson MK, Jacobson BL, Ame J, Lin W;  
XX WPI; 2002-535641/57.  
XX  
XX New nucleic acid molecule encoding bovine poly adenosine diphosphate-  
PT ribose glycohydrolase involved in cellular response to DNA damage,  
PT inhibition of which is useful for treating neoplastic disorders and  
PT neurodegenerative diseases.  
XX  
XX Claim 10; Col 81-82; 77pp; English.  
PS  
XX The invention relates to an isolated nucleic acid molecule which encodes  
CC

CC a bovine poly (adenosine diphosphate (ADP)-ribose) glycohydrolase (bPARG)  
 CC which catalyses release of ADP-ribose from an ADP-ribose polymer. The  
 CC invention is useful as probes and primer molecules that can be used in  
 CC hybridisation assays and polymerase chain reaction (PCR) amplification.  
 CC The knowledge of the nucleotide sequence of the PARG gene permits the  
 CC preparation of antisense therapeutics containing sequences complementary  
 CC to the mRNA of PARG gene. The antisense therapeutic are useful to treat  
 CC neoplastic disorders and conditions caused by genotoxic oxidative stress  
 CC e.g., cardiac disorders, neuronal disorders, reperfusion injury,  
 CC neurotoxicity, Alzheimer's disease, Huntington's disease and Parkinson's  
 CC disease. The invention is useful in gene therapy and antisense therapy.  
 CC The present sequence is bPARG oligopeptide  
 CC  
 CC Sequence 16 AA;

Query Match 94.4%; Score 67; DB 5; Length 16;  
 Best Local Similarity 100.0%; Pred. No. 9.6e-05;  
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IALXLPNIXTQPIPLL 16  
 DB 1 IALXLPNIXTQPIPLL 16

RESULT 2  
 AAU76028  
 ID AAU76028 standard; peptide; 16 AA.

XX AAU76028;  
 AC  
 DT 08-MAY-2002 (first entry)

DE Bovine poly (ADP-ribose) glycohydrolase (PARG) oligopeptide 75.

XX Bovine; poly (ADP-ribose) glycohydrolase; PARG; DNA repair; ADP-ribose;  
 KW adenosine diphosphate-ribose; neoplastic disorder; adenocarcinoma;  
 KW leukaemia; lymphoma; melanoma; myeloma; sarcoma; teratocarcinoma;  
 KW hyperplasia; hypertrophy; reperfusion; ischaemia; heart attack; stroke;  
 KW neurodegenerative disease; neurological disorder; Alzheimer's disease;  
 KW Huntington's disease; Parkinson's disease; oligopeptide 75.

XX Bos taurus.

XX Key Location/Qualifiers  
 FH Misc-difference 4 /label= Unknown  
 FT /note= "Xaa is not further defined in the specification"  
 FT Misc-difference 9 /label= Unknown  
 FT /note= "Xaa is not further defined in the specification"

XX US6337202-B1.

XX 08-JAN-2002.

XX 23-FEB-2000; 2000US-00511477.

XX 01-MAY-1998; 98US-0083768P.

XX 30-APR-1999; 99US-00302812.

XX (KENT ) UNIV KENTUCKY RES FOUND.

XX Jacobson MK, Jacobson EL, Ame J, Lin W;

XX WPI; 2002-163240/21.

XX Novel isolated and purified poly (ADP-ribose) glycohydrolase protein  
 PT which catalyses release of ADP-ribose from ADP-ribose polymer, useful for  
 PT treating neoplastic and neurological disorders, heart attack and stroke.

XX Claim 4; Col 25; 81pp; English.

XX The present invention relates to a new poly (ADP-ribose) glycohydrolase

CC (PARG) protein which catalyses release of ADP-ribose from an ADP  
 CC (adenosine diphosphate)-ribose polymer. The PARG molecule of the  
 CC invention is useful for generating antibodies and can be inhibited or  
 CC activated for diagnosing and treating neoplastic disorders such as  
 CC adenocarcinoma, leukaemia, lymphoma, melanoma, myeloma, sarcoma,  
 CC teratocarcinoma, hyperplasia and hypertrophy, reperfusion following  
 CC ischaemia, heart attack, stroke, neurodegenerative diseases, neurological  
 CC disorders including Alzheimer's, Huntington's and Parkinson's diseases,  
 CC and related conditions. PARG levels may be enhanced to suppress DNA  
 CC repair and increase the cell's susceptibility to chemotherapy drugs.  
 CC Antagonists of PARG are administered to treat or prevent neoplastic  
 CC disorders. The present amino acid sequence represents bovine PARG  
 CC oligonucleotide 75. This peptide is one of several PARG oligopeptides  
 CC (AAU76025-AAU76028) of the invention  
 CC  
 CC Sequence 16 AA;

Query Match 94.4%; Score 67; DB 5; Length 16;  
 Best Local Similarity 100.0%; Pred. No. 9.6e-05;  
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IALXLPNIXTQPIPLL 16  
 DB 1 IALXLPNIXTQPIPLL 16

RESULT 3  
 AAU76019  
 ID AAU76019 standard; peptide; 16 AA.

XX AAU76019;

XX 08-MAY-2002 (first entry)

DE Bovine poly (ADP-ribose) glycohydrolase (PARG) oligopeptide 75.

XX Cow; poly (ADP-ribose) glycohydrolase; PARG; PARG modulator; ADP-ribose;  
 KW adenosine diphosphate-ribose; DNA repair; apoptosis; neoplasia; cancer;  
 KW inherited genetic disease; myocardial infarction; vascular stroke; aging;  
 KW neurodegeneration; Huntington's disease; Parkinson's disease;  
 KW Alzheimer's disease; neurotoxicity; oligopeptide 75.

XX Bos taurus.

XX Key Location/Qualifiers  
 FH Misc-difference 4 /label= Unknown  
 FT /note= "Xaa is not further defined in the specification"  
 FT Misc-difference 9 /label= Unknown  
 FT /note= "Xaa is not further defined in the specification"

XX US6333148-B1.

XX 25-DEC-2001.

XX 30-APR-1999; 99US-00302812.

XX 01-MAY-1998; 98US-0083768P.

XX (KENT ) UNIV KENTUCKY RES FOUND.

XX Jacobson MK, Jacobson EL, Ame J, Lin W;

XX WPI; 2002-153820/20.

XX Screening compounds for modulation of poly (ADP-ribose) glycohydrolase,  
 PT useful potentially for treating diseases associated with DNA damage, e.g.  
 PT cancer.

XX Claim 4; Col 25; 80pp; English.

XX The present invention relates to a new method for screening compounds for

ability to modulate activity of an enzyme that hydrolyses ADP (adenosine diphosphate)-ribose from an ADP-ribose polymer. The compounds are inhibitors or activators of PARP (poly(ADP-ribose) glycohydrolase) and are used to treat or prevent any condition associated with DNA damage e.g. neoplasia, inherited genetic diseases, myocardial infarction, vascular stroke, aging and neurodegeneration e.g. Huntington's, Parkinson's or Alzheimer's diseases, or neurotoxicity generally. Compounds identified by the new method are more effective than known inhibitors and have fewer side effects. The present amino acid sequence represents bovine PARP oligopeptide 75. This peptide is one of several PARP oligopeptides (AAU76016-AAU76019) of the invention

Example 2; Page 14; 86pp; English.

The present invention relates to the isolation of poly(ADP-ribose) glycohydrolase (PARG) from several species, and the polynucleotide sequences encoding them. Methods for inhibiting PARG expression or overexpressing PARG are also disclosed. PARG is involved in the cellular response to DNA damage, and is associated with the body's response to neoplastic disorder inducing agents and oxidative stress. The polynucleotide sequences encoding PARG and PARG modulators are useful for preventing, treating, or ameliorating diseases such as neoplastic disorders, myocardial infarction, vascular stroke, neurodegenerative disorders (e.g. Alzheimer's disease, Parkinson's disease or Huntington's disease), inborn genetic errors, reperfusion following ischaemia, aging, and neurotoxicity. The polynucleotide sequences are also useful in gene therapy. The methods are useful for identifying an agent that modulates PARG activity, identifying a mutant PARG allele in an individual, or screening candidate molecules for PARG modulating activity. ABG72283-ABG72286 represent oligopeptides derived from bovine PARG enzyme. The oligopeptides are used to construct degenerate PCR primers for the isolation of cDNA encoding bovine PARG

Sequence 16 AA;

CC The invention relates to an isolated nucleic acid molecule which encodes  
 CC a bovine poly (adenosine diphosphate (ADP)-ribose) glycohydrolase (bPARG)  
 CC which catalyses release of ADP-ribose from an ADP ribose polymer. The  
 CC invention is useful as probes and primer molecules that can be used in  
 CC hybridisation assays and polymerase chain reaction (PCR) amplification.  
 CC The knowledge of the nucleotide sequence of the PARG gene permits the  
 CC preparation of antisense therapeutics containing sequences complementary  
 CC to the mRNA of PARG gene. The antisense therapeutic are useful to treat  
 CC neoplastic disorders and conditions caused by genotoxic oxidative stress  
 CC e.g., cardiac disorders, neuronal disorders, reperfusion injury,  
 CC neurotoxicity, Alzheimer's disease, Huntington's disease and Parkinson's  
 CC disease. The invention is useful in gene therapy and antisense therapy.  
 CC The present sequence is bPARG oligopeptide

XX Sequence 17 AA;

Query Match 91.5%; Score 65; DB 5; Length 17;  
 Best Local Similarity 87.5%; Pred. No. 0.00023;  
 Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 IALXLPNIXTQPIPLL 16  
 ||| ||||| |||||  
 DB 1 IALCLPNICTQPIPLL 16

RESULT 6  
 AAE25631  
 ID AAE25631 standard; protein; 968 AA.

AC AAE25631;  
 DT 04-NOV-2002 (first entry)

DE Murine poly adenosine diphosphate-ribose glycohydrolase (PARG).

XX Murine; poly adenosine diphosphate-ribose glycohydrolase; PARG; enzyme;  
 KW ADP-ribose; neoplastic disorder; genotoxic oxidative stress; neurotoxic;  
 KW cardiac disorder; neuronal disorder; reperfusion injury; neurotoxicity;  
 KW Alzheimer's disease; Huntington's disease; Parkinson's disease; cardiant;  
 KW cytostatic; vasotropic; neuroprotective; anticonvulsant; gene therapy;  
 KW antisense therapy.

XX Mus musculus.

XX US6395543-B1.

XX 28-MAY-2002.

XX 23-FEB-2000; 2000US-00511507.

XX 01-MAY-1998; 98US-0083768P.

XX 30-APR-1999; 99US-00302812.

XX (KENT ) UNIV KENTUCKY RES FOUND.

XX Jacobson MK, Jacobson EL, Ame J, Lin W;

XX WPI; 2002-535641/57.

XX N-PSDB; AAD42083.

XX New nucleic acid molecule encoding bovine poly adenosine diphosphate-  
 PT ribose glycohydrolase involved in cellular response to DNA damage,  
 PT inhibition of which is useful for treating neoplastic disorders and  
 PT neurodegenerative diseases.

XX Claim 3; Col 63-68; 77pp; English.

XX The invention relates to an isolated nucleic acid molecule which encodes  
 CC a bovine poly (adenosine diphosphate (ADP)-ribose) glycohydrolase (bPARG)  
 CC which catalyses release of ADP-ribose from an ADP ribose polymer. The  
 CC invention is useful as probes and primer molecules that can be used in  
 CC hybridisation assays and polymerase chain reaction (PCR) amplification.  
 CC The knowledge of the nucleotide sequence of the PARG gene permits the

CC preparation of antisense therapeutics containing sequences complementary  
 CC to the mRNA of PARG gene. The antisense therapeutic are useful to treat  
 CC neoplastic disorders and conditions caused by genotoxic oxidative stress  
 CC e.g., cardiac disorders, neuronal disorders, reperfusion injury,  
 CC neurotoxicity, Alzheimer's disease, Huntington's disease and Parkinson's  
 CC disease. The invention is useful in gene therapy and antisense therapy.  
 CC The present sequence is murine PARG

XX Sequence 968 AA;

Query Match 91.5%; Score 65; DB 5; Length 968;  
 Best Local Similarity 87.5%; Pred. No. 0.018;  
 Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 IALXLPNIXTQPIPLL 16  
 ||| ||||| |||||  
 DB 592 IALCLPNICTQPIPLL 607

RESULT 7  
 AAU76022  
 ID AAU76022 standard; protein; 968 AA.

XX AAU76022;

XX 08-MAY-2002 (first entry)

DE Mouse poly (ADP-ribose) glycohydrolase (PARG) protein sequence.

XX Mouse; poly (ADP-ribose) glycohydrolase; PARG; DNA repair; ADP-ribose;  
 KW adenosine diphosphate-ribose; neoplastic disorder; adenocarcinoma;  
 KW leukaemia; lymphoma; melanoma; myeloma; sarcoma; teratocarcinoma;  
 KW hyperplasia; hypertrophy; reperfusion; ischaemia; heart attack; stroke;  
 KW neurodegenerative disease; neurological disorder; Alzheimer's disease;  
 KW Huntington's disease; Parkinson's disease.

XX Mus musculus.

XX US6337202-B1.

XX 08-JAN-2002.

XX 23-FEB-2000; 2000US-00511477.

XX 01-MAY-1998; 98US-0083768P.

XX 30-APR-1999; 99US-00302812.

XX (KENT ) UNIV KENTUCKY RES FOUND.

XX Jacobson MK, Jacobson EL, Ame J, Lin W;

XX WPI; 2002-163240/21.

XX N-PSDB; ABK14933.

XX Novel isolated and purified poly (ADP-ribose) glycohydrolase protein  
 PT which catalyses release of ADP-ribose from ADP ribose polymer, useful for  
 PT treating neoplastic and neurological disorders, heart attack and stroke.

XX Claim 2; Col 63-70; 81pp; English.

XX The present invention relates to a new poly (ADP-ribose) glycohydrolase  
 CC (PARG) protein which catalyses release of ADP-ribose from an ADP  
 CC (adenosine diphosphate)-ribose polymer. The PARG molecule of the  
 CC invention is useful for generating antibodies and can be inhibited or  
 CC activated for diagnosing and treating neoplastic disorders such as  
 CC adenocarcinoma, leukaemia, lymphoma, melanoma, myeloma, sarcoma,  
 CC teratocarcinoma, hyperplasia and hypertrophy, reperfusion following  
 CC ischaemia, heart attack, stroke, neurodegenerative diseases, neurological  
 CC disorders including Alzheimer's, Huntington's and Parkinson's diseases,  
 CC and related conditions. PARG levels may be enhanced to suppress DNA  
 CC repair and increase the cell's susceptibility to chemotherapy drugs.  
 CC Antagonists of PARG are administered to treat or prevent neoplastic  
 CC disorders. The present amino acid sequence represents the mouse PARG

Best Local Similarity 87.5%; Pred. No. 0.018;  
Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Best Local Similarity

Query Match 91.5%; Score 65; DB 6; Length 968;  
Best Local Similarity 87.5%; Pred. No. 0.018;







us-09-302-812-14.rag

Thu May 27 09:55:59 2004

KW ADP-ribose; neoplastic disorder; genotoxic oxidative stress; neutropenic;  
 KW cardiac disorder; neuronal disorder; reperfusion injury; neurotoxicity;  
 KW Alzheimer's disease; Huntington's disease; Parkinson's disease; cardiac;  
 KW cytosolic; vasotropic; neuroprotective; anticonvulsant; gene therapy;  
 KW antisense therapy.

XX Bos taurus.

OS US6395543-B1.

PN 28-MAY-2002.

XX 23-FEB-2000; 2000US-00511507.

XX 01-MAY-1998; 98US-0083768P.

XX 30-APR-1999; 99US-00302812.

XX (KENT ) UNIV KENTUCKY RES FOUND.

XX Jacobson MK, Jacobson EL, Ame J, Lin W;

XX WPI; 2002-535641/57.

XX N-PSDB; AAD42081.

XX New nucleic acid molecule encoding bovine poly adenosine diphosphate-  
 ribose glycohydrolase involved in cellular response to DNA damage, and  
 inhibition of which is useful for treating neoplastic disorders and  
 neurodegenerative diseases.

XX Claim 3; Col 47-45; 77pp; English.

XX The invention relates to an isolated nucleic acid molecule which encodes  
 a bovine poly (adenosine diphosphate (ADP)-ribose) glycohydrolase (bPARG)  
 which catalyses release of ADP-ribose from an ADP ribose polymer. The  
 invention is useful as probes and primer molecules that can be used in  
 hybridisation assays and polymerase chain reaction (PCR) amplification.  
 The knowledge of the nucleotide sequence of the PARG gene permits the  
 preparation of antisense therapeutics containing sequences complementary  
 to the mRNA of PARG gene. The antisense therapeutic are useful to treat  
 neoplastic disorders and conditions caused by genotoxic oxidative stress  
 e.g., cardiac disorders, neuronal disorders, reperfusion injury,  
 neurotoxicity, Alzheimer's disease, Huntington's disease and Parkinson's  
 disease. The invention is useful in gene therapy and antisense therapy.  
 The present sequence is bPARG

XX Sequence 977 AA;

Query Match 91.5%; Score 65; DB 5; Length 977;

Best Local Similarity 87.5%; Pred. No. 0.018;

Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 IALXLPNIXTQPIPL 16

Db 601 IALCLPNICTQPIPL 616

RESULT 15

AAU76020

XX AAU76020 standard; protein; 977 AA.

XX AAU76020;

XX 08-MAY-2002 (first entry)

XX Bovine poly(ADP-ribose) glycohydrolase (PARG) protein sequence.

XX Cow; poly(ADP-ribose) glycohydrolase; PARG; DNA repair; ADP-ribose;  
 KW adenosine diphosphate-ribose; neoplastic disorder; adenocarcinoma;  
 KW leukaemia; lymphoma; melanoma; myeloma; sarcoma; teratocarcinoma;  
 KW hyperplasia; hypertrophy; reperfusion; ischaemia; heart attack; stroke;  
 KW neurodegenerative disease; neurological disorder; Alzheimer's disease;  
 KW Huntington's disease; Parkinson's disease.

OS Bos taurus.

XX Key

XX Region

XX Region

XX Region

XX Region

XX Region

XX Region

XX Region

XX Region

XX Region

XX Region

XX Region

XX Region

XX Region

XX Region

XX Region

XX Region

XX Region

XX Region

XX Region

XX Region

XX Region

XX Region

XX Region

XX Region

XX Region

XX Region

XX Region

XX Region

XX Region

XX Region

XX Region

XX Region

XX Region

XX Region

XX Region

XX Region

XX Region

Location/Qualifiers

601. 617

/note= "Represents PARG oligopeptide #1"

761. 770

/note= "Represents PARG oligopeptide #2"

771. 801

/note= "Represents PARG oligopeptide #3"

849. 880

/note= "Represents PARG oligopeptide #4"

US6337202-B1.

08-JAN-2002.

23-FEB-2000; 2000US-00511477.

01-MAY-1998; 98US-0083768P.

30-APR-1999; 99US-00302812.

(KENT ) UNIV KENTUCKY RES FOUND.

Jacobson MK, Jacobson EL, Ame J, Lin W;

WPI; 2002-163240/21.

N-PSDB; ABK14931.

Novel isolated and purified poly (ADP-ribose) glycohydrolase protein  
 which catalyses release of ADP-ribose from ADP ribose polymer, useful for  
 treating neoplastic and neurological disorders, heart attack and stroke.

Claim 2; Col 47-52; 81pp; English.

The present invention relates to a new poly(ADP-ribose) glycohydrolase  
 (PARG) protein which catalyses release of ADP-ribose from an ADP  
 (adenosine diphosphate)-ribose polymer. The PARG molecule of the  
 invention is useful for generating antibodies and can be inhibited or  
 activated for diagnosing and treating neoplastic disorders such as  
 adenocarcinoma, leukaemia, lymphoma, melanoma, myeloma, sarcoma,  
 teratocarcinoma, hyperplasia and hypertrophy, reperfusion following  
 ischaemia, heart attack, stroke, neurodegenerative diseases, neurological  
 disorders including Alzheimer's, Huntington's and Parkinson's diseases,  
 and related conditions. PARG levels may be enhanced to suppress DNA  
 repair and increase the cell's susceptibility to chemotherapy drugs.  
 Antagonists of PARG are administered to treat or prevent neoplastic  
 disorders. The present amino acid sequence represents the bovine PARG  
 protein of the invention. This protein is one of several PARG proteins  
 (AAU76020-AAU76024) of the invention

Sequence 977 AA;

Query Match 91.5%; Score 65; DB 5; Length 977;

Best Local Similarity 87.5%; Pred. No. 0.018;

Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 IALXLPNIXTQPIPL 16

Db 601 IALCLPNICTQPIPL 616

Search completed: May 26, 2004, 18:40:13

Job time : 2.83313 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 26, 2004, 18:38:05 ; Search time 0.512697 Seconds  
(without alignments)  
1611.120 Million cell updates/sec

Title: US-09-302-812-14  
Perfect score: 71 IALXLPNIXTQPIPLL 16  
Sequence:

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents AA:\*  
1: /cgn2\_6/ptodata/2/iaa/5A COMB.pep.\*  
2: /cgn2\_6/ptodata/2/iaa/5B COMB.pep.\*  
3: /cgn2\_6/ptodata/2/iaa/6A COMB.pep.\*  
4: /cgn2\_6/ptodata/2/iaa/6B COMB.pep.\*  
5: /cgn2\_6/ptodata/2/iaa/PCTUS COMB.pep.\*  
6: /cgn2\_6/ptodata/2/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	67	94.4	16	4	US-09-302-812-14
2	67	94.4	16	4	Sequence 14, Appl
3	67	94.4	16	4	Sequence 14, Appl
4	65	91.5	968	4	US-09-511-477-14
5	65	91.5	968	4	Sequence 6, Appl
6	65	91.5	968	4	Sequence 6, Appl
7	65	91.5	968	4	Sequence 6, Appl
8	65	91.5	968	4	Sequence 6, Appl
9	65	91.5	968	4	Sequence 6, Appl
10	65	91.5	968	4	Sequence 6, Appl
11	65	91.5	968	4	Sequence 6, Appl
12	65	91.5	968	4	Sequence 6, Appl
13	46	64.8	768	4	US-09-302-812-8
14	46	64.8	768	4	Sequence 8, Appl
15	46	64.8	768	4	Sequence 8, Appl
16	40	56.3	160	4	US-09-511-507-8
17	40	56.3	373	4	US-09-489-039A-11236
18	38	53.5	342	4	US-09-254-077A-12
19	38	53.5	453	1	US-09-543-681A-5179
20	38	53.5	453	1	US-08-439-131A-5
21	38	53.5	453	1	US-08-440-674-4
22	38	53.5	453	4	US-08-879-337-6
23	37	52.1	181	4	US-09-543-681A-4556
24	37	52.1	341	4	Sequence 3607, Ap
25	37	52.1	366	4	US-09-489-039A-8294
26	37	52.1	415	4	US-09-540-236-3007
27	37	52.1	437	4	US-09-543-681A-6746
					Sequence 12063, A

28	37	52.1	542	4	US-09-614-891-10	Sequence 10, Appl
29	37	52.1	584	4	US-09-489-039A-14137	Sequence 14137, A
30	37	52.1	597	1	US-08-399-696-102	Sequence 102, App
31	36	50.7	233	4	US-09-489-039A-12718	Sequence 12718, A
32	36	50.7	331	4	US-09-454-279-12	Sequence 12, Appl
33	36	50.7	341	4	US-09-543-681A-4713	Sequence 4713, Ap
34	36	50.7	445	4	US-09-252-991A-22769	Sequence 22769, A
35	36	50.7	566	4	US-09-252-991A-17972	Sequence 17972, A
36	35	49.3	101	4	US-09-345-236B-111	Sequence 111, App
37	35	49.3	425	4	US-09-634-955B-19	Sequence 19, Appl
38	35	49.3	466	4	US-09-489-039A-12781	Sequence 12781, A
39	35	49.3	583	4	US-09-107-532A-5678	Sequence 5678, Ap
40	35	49.3	689	4	US-09-252-991A-31790	Sequence 31790, A
41	34	47.9	16	3	US-09-179-558-12	Sequence 12, Appl
42	34	47.9	16	4	US-09-722-825-12	Sequence 12, Appl
43	34	47.9	16	4	US-09-722-487-12	Sequence 12, Appl
44	34	47.9	16	4	US-09-722-708-12	Sequence 12, Appl
45	34	47.9	25	1	US-08-190-130-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1

US-09-302-812-14

; Sequence 14, Application US/09302812B

; Patent No. 6333148

; GENERAL INFORMATION:

; APPLICANT: JACOBSON, Myron K.

; APPLICANT: JACOBSON, Elaine L.

; APPLICANT: AM, Jean-Christophe

; APPLICANT: LIN, Winston

; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE (PARG) ENZ

; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIVE

; TITLE OF INVENTION: THEREWITH

; FILE REFERENCE: NIAD 201

; CURRENT APPLICATION NUMBER: US/09/302,812B

; CURRENT FILING DATE: 1999-04-30

; EARLIER APPLICATION NUMBER: 60/083,768

; EARLIER FILING DATE: 1998-05-01

; NUMBER OF SEQ ID NOS: 38

; SEQ ID NO 14

; LENGTH: 16

; TYPE: PRT

; ORGANISM: Bos taurus

; FEATURE:

US-09-302-812-14

Query Match 94.4%; Score 67; DB 4; Length 16;

Best Local Similarity 100.0%; Pred. No. 4.8e-05;

Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 IALXLPNIXTQPIPLL 16

Db 1 IALXLPNIXTQPIPLL 16

RESULT 2

US-09-511-477-14

; Sequence 14, Application US/09511477

; Patent No. 6337202

; GENERAL INFORMATION:

; APPLICANT: JACOBSON, Myron K.

; APPLICANT: JACOBSON, Elaine L.

; APPLICANT: AME, Jean-Christophe

; APPLICANT: LIN, Winston

; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE (PARG) ENZ

; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIVE

; TITLE OF INVENTION: THEREWITH

; FILE REFERENCE: NIAD 201

; CURRENT APPLICATION NUMBER: US/09/511,477

; CURRENT FILING DATE: 2000-02-23

; PRIOR APPLICATION NUMBER: 09/302,812

```

; PRIOR FILING DATE: 1999-04-30
; NUMBER OF SEQ ID NOS: 38
; SEQ ID NO 14
; LENGTH: 16
; TYPE: PRT
; ORGANISM: Bos taurus
; FEATURE:
; US-09-511-477-14

Query Match          94.4%; Score 67; DB 4; Length 16;
Best Local Similarity 100.0%; Pred. No. 4.8e-05;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IALXLPNIXTOPIPLL 16
Db 1 IALXLPNIXTOPIPLL 16

RESULT 3
US-09-511-507-14
; Sequence 14, Application US/09511507
; Patent No. 6395543
; GENERAL INFORMATION:
; APPLICANT: JACOBSON, Myron K.
; APPLICANT: JACOBSON, Elaine L.
; APPLICANT: AME, Jean-Christophe
; APPLICANT: LIN, Winston
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE (PARG) EN
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV
; TITLE OF INVENTION: THEREWITH
; FILE REFERENCE: NIAD 201
; CURRENT APPLICATION NUMBER: US/09/511,507
; CURRENT FILING DATE: 2000-02-23
; PRIOR APPLICATION NUMBER: 09/302,812
; PRIOR FILING DATE: 1999-04-30
; NUMBER OF SEQ ID NOS: 38
; SEQ ID NO 14
; LENGTH: 16
; TYPE: PRT
; ORGANISM: Bos taurus
; FEATURE:
; US-09-511-507-14

Query Match          94.4%; Score 67; DB 4; Length 16;
Best Local Similarity 100.0%; Pred. No. 4.8e-05;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IALXLPNIXTOPIPLL 16
Db 1 IALXLPNIXTOPIPLL 16

RESULT 4
US-09-302-812-6
; Sequence 6, Application US/09302812B
; Patent No. 6333148
; GENERAL INFORMATION:
; APPLICANT: JACOBSON, Myron K.
; APPLICANT: JACOBSON, Elaine L.
; APPLICANT: AME, Jean-Christophe
; APPLICANT: LIN, Winston
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE (PARG) ENZ
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV
; TITLE OF INVENTION: THEREWITH
; FILE REFERENCE: NIAD 201
; CURRENT APPLICATION NUMBER: US/09/302,812B
; CURRENT FILING DATE: 1999-04-30
; EARLIER APPLICATION NUMBER: 60/083,768
; EARLIER FILING DATE: 1998-05-01
; NUMBER OF SEQ ID NOS: 38
; SEQ ID NO 6
; LENGTH: 968
; TYPE: PRT

```

```

; ORGANISM: Mus musculus
; FEATURE:
; US-09-302-812-6

Query Match          91.5%; Score 65; DB 4; Length 968;
Best Local Similarity 87.5%; Pred. No. 0.0095;
Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 IALXLPNIXTOPIPLL 16
Db 592 IALCLPNICTOPIPLL 607

RESULT 5
US-09-511-477-6
; Sequence 6, Application US/09511477
; Patent No. 6337202
; GENERAL INFORMATION:
; APPLICANT: JACOBSON, Myron K.
; APPLICANT: JACOBSON, Elaine L.
; APPLICANT: AME, Jean-Christophe
; APPLICANT: LIN, Winston
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE (PARG) EN
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV
; TITLE OF INVENTION: THEREWITH
; FILE REFERENCE: NIAD 201
; CURRENT APPLICATION NUMBER: US/09/511,477
; CURRENT FILING DATE: 2000-02-23
; PRIOR APPLICATION NUMBER: 09/302,812
; PRIOR FILING DATE: 1999-04-30
; NUMBER OF SEQ ID NOS: 38
; SEQ ID NO 6
; LENGTH: 968
; TYPE: PRT
; ORGANISM: Mus musculus
; FEATURE:
; US-09-511-477-6

Query Match          91.5%; Score 65; DB 4; Length 968;
Best Local Similarity 87.5%; Pred. No. 0.0095;
Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 IALXLPNIXTOPIPLL 16
Db 592 IALCLPNICTOPIPLL 607

RESULT 6
US-09-511-507-6
; Sequence 6, Application US/09511507
; Patent No. 6395543
; GENERAL INFORMATION:
; APPLICANT: JACOBSON, Myron K.
; APPLICANT: JACOBSON, Elaine L.
; APPLICANT: AME, Jean-Christophe
; APPLICANT: LIN, Winston
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE (PARG) EN
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV
; TITLE OF INVENTION: THEREWITH
; FILE REFERENCE: NIAD 201
; CURRENT APPLICATION NUMBER: US/09/511,507
; CURRENT FILING DATE: 2000-02-23
; PRIOR APPLICATION NUMBER: 09/302,812
; PRIOR FILING DATE: 1999-04-30
; NUMBER OF SEQ ID NOS: 38
; SEQ ID NO 6
; LENGTH: 968
; TYPE: PRT
; ORGANISM: Mus musculus
; FEATURE:
; US-09-511-507-6

Query Match          91.5%; Score 65; DB 4; Length 968;

```

```
DB      600 IALCLPNICTQPIPLL 615

RESULT 9
US-09-511-507-4
; Sequence 4, Application US/095111507
; Patent No. 6395543
; GENERAL INFORMATION:
; APPLICANT: JACOBSON, Myron K.
; APPLICANT: JACOBSON, Elaine L.
; APPLICANT: AME, Jean-Christophe
; APPLICANT: LIN, Winston
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOSYLTRANSFERASE (PARG) ENZYME
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIVE
; TITLE OF INVENTION: THEREWITH
; FILE REFERENCE: NIAD 201
; CURRENT APPLICATION NUMBER: US/09/511,507
; CURRENT FILING DATE: 2000-02-23
; PRIOR APPLICATION NUMBER: 09/302,812
; PRIOR FILING DATE: 1999-04-30
; NUMBER OF SEQ ID NOS: 38
; SEQ ID NO 4
; LENGTH: 976
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
;
US-09-511-507-4

Query Match          91.5%; Score 65; DB 4; Length 976;
Best Local Similarity 87.5%; Pred. No. 0.0095; 2; Indels 0; Gaps 0;
Matches 14; Conservative 0; Mismatches 2;

QY      1 IALXLPNIXTQPIPLL 16
      ||| |||| |||| |||| ||||
DB      600 IALCLPNICTQPIPLL 615


RESULT 10
US-09-302-812-2
; Sequence 2, Application US/09302812B
; Patent No. 6333148
; GENERAL INFORMATION:
; APPLICANT: JACOBSON, Myron K.
; APPLICANT: JACOBSON, Elaine L.
; APPLICANT: AM, Jean-Christophe
; APPLICANT: LIN, Winston
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOSYLTRANSFERASE (PARG) ENZYME
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIVE
; TITLE OF INVENTION: THEREWITH
; FILE REFERENCE: NIAD 201
; CURRENT APPLICATION NUMBER: US/09/302,812B
; CURRENT FILING DATE: 1999-04-30
; EARLIER APPLICATION NUMBER: 60/083,768
; EARLIER FILING DATE: 1998-05-01
; NUMBER OF SEQ ID NOS: 38
; SEQ ID NO 2
; LENGTH: 977
; TYPE: PRT
; ORGANISM: Bos taurus
; FEATURE:
;
US-09-302-812-2

Query Match          91.5%; Score 65; DB 4; Length 977;
Best Local Similarity 87.5%; Pred. No. 0.0096; 2; Indels 0; Gaps 0;
Matches 14; Conservative 0; Mismatches 2;

QY      1 IALXLPNIXTQPIPLL 16
      ||| |||| |||| |||| ||||
DB      601 IALCLPNICTQPIPLL 616


RESULT 11
US-09-511-477-2
```

```

; Sequence 2, Application US/09511477
; Patent No. 6337202
; GENERAL INFORMATION:
; APPLICANT: JACOBSON, Myron K.
; APPLICANT: JACOBSON, Elaine L.
; APPLICANT: AME, Jean-Christophe
; APPLICANT: LIN, Winston
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE (PARG) EN
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV
; TITLE OF INVENTION: THEREWITH
; FILE REFERENCE: NIAD 201
; CURRENT APPLICATION NUMBER: US/09/511,477
; CURRENT FILING DATE: 2000-02-23
; PRIOR APPLICATION NUMBER: 09/302,812
; PRIOR FILING DATE: 1999-04-30
; NUMBER OF SEQ ID NOS: 38
; SEQ ID NO 2
; LENGTH: 977
; TYPE: PRT
; ORGANISM: Bos taurus
; FEATURE:
US-09-511-477-2

Query Match 91.5%; Score 65; DB 4; Length 977;
Best Local Similarity 87.5%; Pred. No. 0.0096;
Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 IALXLPNIXTOPIPL 16
Db 601 IALCLPNICTOPIPL 616

RESULT 12
US-09-511-507-2
; Sequence 2, Application US/09511507
; Patent No. 6395543
; GENERAL INFORMATION:
; APPLICANT: JACOBSON, Myron K.
; APPLICANT: JACOBSON, Elaine L.
; APPLICANT: AME, Jean-Christophe
; APPLICANT: LIN, Winston
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE (PARG) EN
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV
; TITLE OF INVENTION: THEREWITH
; FILE REFERENCE: NIAD 201
; CURRENT APPLICATION NUMBER: US/09/511,507
; CURRENT FILING DATE: 2000-02-23
; PRIOR APPLICATION NUMBER: 09/302,812
; PRIOR FILING DATE: 1999-04-30
; NUMBER OF SEQ ID NOS: 38
; SEQ ID NO 2
; LENGTH: 977
; TYPE: PRT
; ORGANISM: Bos taurus
; FEATURE:
US-09-511-507-2

Query Match 91.5%; Score 65; DB 4; Length 977;
Best Local Similarity 87.5%; Pred. No. 0.0096;
Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 IALXLPNIXTOPIPL 16
Db 601 IALCLPNICTOPIPL 616

RESULT 13
US-09-302-812-8
; Sequence 8, Application US/09302812B
; Patent No. 6333148
; GENERAL INFORMATION:
; APPLICANT: JACOBSON, Myron K.
; APPLICANT: JACOBSON, Elaine L.

```

```

; APPLICANT: AM, Jean-Christophe
; APPLICANT: LIN, Winston
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE (PARG) ENZ
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIVE
; TITLE OF INVENTION: THEREWITH
; FILE REFERENCE: NIAD 201
; CURRENT APPLICATION NUMBER: US/09/302,812B
; CURRENT FILING DATE: 1999-04-30
; EARLIER APPLICATION NUMBER: 60/083,768
; EARLIER FILING DATE: 1998-05-01
; NUMBER OF SEQ ID NOS: 38
; SEQ ID NO 8
; LENGTH: 768
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
; FEATURE:
US-09-302-812-8

Query Match 64.8%; Score 46; DB 4; Length 768;
Best Local Similarity 50.0%; Pred. No. 9.8;
Matches 8; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 1 IALXLPNIXTOPIPL 16
Db 219 LALRLPDIQSPVPL 234

RESULT 14
US-09-511-477-8
; Sequence 8, Application US/09511477
; Patent No. 6337202
; GENERAL INFORMATION:
; APPLICANT: JACOBSON, Myron K.
; APPLICANT: JACOBSON, Elaine L.
; APPLICANT: AME, Jean-Christophe
; APPLICANT: LIN, Winston
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE (PARG) EN
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV
; TITLE OF INVENTION: THEREWITH
; FILE REFERENCE: NIAD 201
; CURRENT APPLICATION NUMBER: US/09/511,477
; CURRENT FILING DATE: 2000-02-23
; PRIOR APPLICATION NUMBER: 09/302,812
; PRIOR FILING DATE: 1999-04-30
; NUMBER OF SEQ ID NOS: 38
; SEQ ID NO 8
; LENGTH: 768
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
; FEATURE:
US-09-511-477-8

Query Match 64.8%; Score 46; DB 4; Length 768;
Best Local Similarity 50.0%; Pred. No. 9.8;
Matches 8; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 1 IALXLPNIXTOPIPL 16
Db 219 LALRLPDIQSPVPL 234

RESULT 15
US-09-511-507-8
; Sequence 8, Application US/09511507
; Patent No. 6395543
; GENERAL INFORMATION:
; APPLICANT: JACOBSON, Myron K.
; APPLICANT: JACOBSON, Elaine L.
; APPLICANT: AME, Jean-Christophe
; APPLICANT: LIN, Winston
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE (PARG) EN
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV
; TITLE OF INVENTION: THEREWITH

```

us-09-302-812-14.ra1

Thu May 27 09:55:59 2004

```

; FILE REFERENCE: NIAD 201
; CURRENT APPLICATION NUMBER: US/09/511,507
; CURRENT FILING DATE: 2000-02-23
; PRIOR APPLICATION NUMBER: 09/302,812
; PRIOR FILING DATE: 1999-04-30
; NUMBER OF SEQ ID NOS: 38
; SEQ ID NO 8
; LENGTH: 768
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
; FEATURE:
US-09-511-507-8

Query Match      64.8%; Score 46; DB 4; Length 768;
Best Local Similarity 50.0%; Pred. No. 9.8;
Matches 8; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 1 LALXLPNIXTQPIPLL 16
   :|||:::|:|
Db 219 LALRLPDLIQSEVPILL 234

Search completed: May 26, 2004, 18:49:32
Job time : 0.512697 secs

```



Thu May 27 09:55:59 2004

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 26, 2004, 18:46:21 ; Search time 1.33495 Seconds

(without alignments)  
3343.018 Million cell updates/sec

Title: US-09-302-812-14

Perfect score: 71

Sequence: 1 IALXLPNIXTOPIPLL 16

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1149313 seqs, 278921704 residues

Total number of hits satisfying chosen parameters: 1149313

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications AA:\*

- 1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*
- 2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pep.\*
- 3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep.\*
- 4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep.\*
- 5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep.\*
- 6: /cgn2\_6/ptodata/1/pubpaa/PCTUS\_PUBCOMB.pep.\*
- 7: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep.\*
- 8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep.\*
- 9: /cgn2\_6/ptodata/1/pubpaa/US09A\_PUBCOMB.pep.\*
- 10: /cgn2\_6/ptodata/1/pubpaa/US09B\_PUBCOMB.pep.\*
- 11: /cgn2\_6/ptodata/1/pubpaa/US09C\_PUBCOMB.pep.\*
- 12: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep.\*
- 13: /cgn2\_6/ptodata/1/pubpaa/US10A\_PUBCOMB.pep.\*
- 14: /cgn2\_6/ptodata/1/pubpaa/US10B\_PUBCOMB.pep.\*
- 15: /cgn2\_6/ptodata/1/pubpaa/US10C\_PUBCOMB.pep.\*
- 16: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep.\*
- 17: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep.\*
- 18: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	67	94.4	16	9	US-09-973-451-14
2	65	91.5	968	9	US-09-973-451-6
3	65	91.5	976	9	US-09-973-451-4
4	65	91.5	977	9	US-09-973-451-2
5	46	64.8	768	9	US-09-973-451-8
6	43	60.6	521	12	US-10-412-699B-82
7	43	60.6	521	12	US-10-225-066A-922
8	43	60.6	521	14	US-10-278-173-36
9	43	60.6	521	14	US-10-278-536-34
10	43	60.6	521	15	US-10-374-780A-2244
11	41	57.7	79	12	US-10-424-599-182136
12	41	57.7	499	14	US-10-156-761-9332
13	41	57.7	1089	15	US-10-369-493-2154
14	41	57.7	1579	9	US-09-801-368-368
15	41	57.7	1579	15	US-10-369-493-2000

16	56.3	429	12	US-10-282-122A-58079	Sequence 58079, A
17	54.9	228	15	US-10-012-697-1529	GENERAL INFORMATION
18	53.5	56	9	US-09-796-692-1133	Sequence 1133, Ap
19	53.5	56	9	US-09-796-692-1583	Sequence 1583, Ap
20	53.5	56	9	US-09-796-692-2200	Sequence 2200, Ap
21	53.5	56	14	US-10-040-862-1133	Sequence 1133, Ap
22	53.5	56	14	US-10-040-862-1583	Sequence 1583, Ap
23	53.5	56	14	US-10-040-862-2200	Sequence 2200, Ap
24	53.5	56	15	US-10-057-475B-1133	Sequence 1133, Ap
25	53.5	56	15	US-10-057-475B-1583	Sequence 1583, Ap
26	53.5	56	15	US-10-057-475B-2200	Sequence 2200, Ap
27	53.5	56	15	US-10-154-884B-1133	Sequence 1133, Ap
28	53.5	56	15	US-10-154-884B-1583	Sequence 1583, Ap
29	53.5	56	15	US-10-154-884B-2200	Sequence 2200, Ap
30	53.5	329	12	US-10-282-122A-60939	Sequence 60939, A
31	53.5	348	9	US-09-730-617-2	Sequence 2, Appli
32	53.5	426	12	US-10-282-122A-61195	Sequence 61195, A
33	53.5	430	12	US-10-282-122A-55560	Sequence 55560, A
34	53.5	453	8	US-08-879-337-6	Sequence 6, Appli
35	53.5	453	15	US-10-369-493-2273	Sequence 2273, Ap
36	53.5	461	15	US-10-108-260A-3238	Sequence 3238, Ap
37	53.5	477	12	US-10-425-114-65173	Sequence 65173, A
38	53.5	616	9	US-09-925-300-1519	Sequence 1519, Ap
39	53.5	651	9	US-09-815-242-11022	Sequence 11022, A
40	53.5	651	12	US-10-282-122A-58169	Sequence 58169, A
41	53.5	702	12	US-10-282-122A-47346	Sequence 47346, A
42	53.5	713	15	US-10-369-493-13792	Sequence 13792, A
43	53.5	766	9	US-09-801-368-316	Sequence 316, App
44	53.5	950	15	US-10-369-493-6705	Sequence 6705, Ap
45	53.5	952	15	US-10-369-493-6704	Sequence 6704, Ap

ALIGNMENTS

RESULT 1

US-09-973-451-14  
; Sequence 14, Application US/09973451  
; Patent No. US20020132328A1  
; GENERAL INFORMATION:  
; APPLICANT: JACOBSON, Myron K.  
; APPLICANT: JACOBSON, Elaine L.  
; APPLICANT: AM, Jean-Christophe  
; APPLICANT: LIN, Winston  
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE  
; TITLE OF INVENTION: (PARG) ENZYMES,  
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV  
; TITLE OF INVENTION: THEREWITH  
; FILE REFERENCE: NIAD 201  
; CURRENT APPLICATION NUMBER: US/09/973,451  
; CURRENT FILING DATE: 2001-10-09  
; PRIOR APPLICATION NUMBER: US/09/302,812  
; PRIOR FILING DATE: 1999-04-30  
; PRIOR APPLICATION NUMBER: 60/083,768  
; PRIOR FILING DATE: 1998-05-01  
; NUMBER OF SEQ ID NOS: 38  
; SEQ ID NO 14  
; LENGTH: 16  
; TYPE: PRT  
; ORGANISM: Bos taurus  
; FEATURE:  
US-09-973-451-14

Query Match 94.4%; Score 67; DB 9; Length 16;  
Best Local Similarity 100.0%; Pred. No. 0.00028;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 IALXLPNIXTOPIPLL 16  
Db 1 IALXLPNIXTOPIPLL 16

RESULT 2

us-09-302-812-14.rapb

Thu May 27 09:55:59 2004

```

US-09-973-451-6
; Sequence 6, Application US/09973451
; Patent No. US20020132328A1
; GENERAL INFORMATION:
; APPLICANT: JACOBSON, Myron K.
; APPLICANT: JACOBSON, Elaine L.
; APPLICANT: AM, Jean-Christophe
; APPLICANT: LIN, Winston
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE
; TITLE OF INVENTION: (PARG) ENZYMES,
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV
; TITLE OF INVENTION: THEREWITH
; FILE REFERENCE: NIAD 201
; CURRENT APPLICATION NUMBER: US/09/973,451
; CURRENT FILING DATE: 2001-10-09
; PRIOR APPLICATION NUMBER: US/09/302,812
; PRIOR FILING DATE: 1999-04-30
; PRIOR APPLICATION NUMBER: 60/083,768
; PRIOR FILING DATE: 1998-05-01
; NUMBER OF SEQ ID NOS: 38
; SEQ ID NO 6
; LENGTH: 968
; TYPE: PRT
; ORGANISM: Mus musculus
; FEATURE:
;
US-09-973-451-6
Query Match 91.5%; Score 65; DB 9; Length 968;
Best Local Similarity 87.5%; Pred. No. 0.046;
Matches 14; Conservative 0; Mismatches 2; Indels 2; Gaps 0;

QY 1 IALXLPNIXTOPIPLL 16
DB 592 IALCLPNICTQPIPLL 607

RESULT 3
US-09-973-451-4
; Sequence 4, Application US/09973451
; Patent No. US20020132328A1
; GENERAL INFORMATION:
; APPLICANT: JACOBSON, Myron K.
; APPLICANT: JACOBSON, Elaine L.
; APPLICANT: AM, Jean-Christophe
; APPLICANT: LIN, Winston
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE
; TITLE OF INVENTION: (PARG) ENZYMES,
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV
; TITLE OF INVENTION: THEREWITH
; FILE REFERENCE: NIAD 201
; CURRENT APPLICATION NUMBER: US/09/973,451
; CURRENT FILING DATE: 2001-10-09
; PRIOR APPLICATION NUMBER: US/09/302,812
; PRIOR FILING DATE: 1999-04-30
; PRIOR APPLICATION NUMBER: 60/083,768
; PRIOR FILING DATE: 1998-05-01
; NUMBER OF SEQ ID NOS: 38
; SEQ ID NO 4
; LENGTH: 976
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
;
US-09-973-451-4
Query Match 91.5%; Score 65; DB 9; Length 976;
Best Local Similarity 87.5%; Pred. No. 0.046;
Matches 14; Conservative 0; Mismatches 2; Indels 2; Gaps 0;

QY 1 IALXLPNIXTOPIPLL 16
DB 600 IALCLPNICTQPIPLL 615

US-09-973-451-2
; Sequence 2, Application US/09973451
; Patent No. US20020132328A1
; GENERAL INFORMATION:
; APPLICANT: JACOBSON, Myron K.
; APPLICANT: JACOBSON, Elaine L.
; APPLICANT: AM, Jean-Christophe
; APPLICANT: LIN, Winston
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE
; TITLE OF INVENTION: (PARG) ENZYMES,
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV
; TITLE OF INVENTION: THEREWITH
; FILE REFERENCE: NIAD 201
; CURRENT APPLICATION NUMBER: US/09/973,451
; CURRENT FILING DATE: 2001-10-09
; PRIOR APPLICATION NUMBER: US/09/302,812
; PRIOR FILING DATE: 1999-04-30
; PRIOR APPLICATION NUMBER: 60/083,768
; PRIOR FILING DATE: 1998-05-01
; NUMBER OF SEQ ID NOS: 38
; SEQ ID NO 2
; LENGTH: 977
; TYPE: PRT
; ORGANISM: Bos taurus
; FEATURE:
;
US-09-973-451-2
Query Match 91.5%; Score 65; DB 9; Length 977;
Best Local Similarity 87.5%; Pred. No. 0.046;
Matches 14; Conservative 0; Mismatches 2; Indels 2; Gaps 0;

QY 1 IALXLPNIXTOPIPLL 16
DB 601 IALCLPNICTQPIPLL 616

RESULT 5
US-09-973-451-8
; Sequence 8, Application US/09973451
; Patent No. US20020132328A1
; GENERAL INFORMATION:
; APPLICANT: JACOBSON, Myron K.
; APPLICANT: JACOBSON, Elaine L.
; APPLICANT: AM, Jean-Christophe
; APPLICANT: LIN, Winston
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE
; TITLE OF INVENTION: (PARG) ENZYMES,
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV
; TITLE OF INVENTION: THEREWITH
; FILE REFERENCE: NIAD 201
; CURRENT APPLICATION NUMBER: US/09/973,451
; CURRENT FILING DATE: 2001-10-09
; PRIOR APPLICATION NUMBER: US/09/302,812
; PRIOR FILING DATE: 1999-04-30
; PRIOR APPLICATION NUMBER: 60/083,768
; PRIOR FILING DATE: 1998-05-01
; NUMBER OF SEQ ID NOS: 38
; SEQ ID NO 8
; LENGTH: 768
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
; FEATURE:
;
US-09-973-451-8
Query Match 64.8%; Score 46; DB 9; Length 768;
Best Local Similarity 50.0%; Pred. No. 43;
Matches 8; Conservative 4; Mismatches 4; Indels 4; Gaps 0;

QY 1 IALXLPNIXTOPIPLL 16
DB 219 IALRLPDLIQSPVPLL 234

```

```

; GENERAL INFORMATION:
; APPLICANT: Mendel Biotechnology, Inc.
; APPLICANT: RATCLIFFE, Oliver
; APPLICANT: RIECHMANN, Jose Luis
; APPLICANT: ADAM, Luc J
; APPLICANT: DUBELL, Arnold T
; APPLICANT: HEARD, Jacqueline E
; APPLICANT: PILGRIM, Marsha L
; APPLICANT: JIANG, Cai-Zhong
; APPLICANT: REUBER, T. Lynne
; APPLICANT: CREELMAN, Robert A
; APPLICANT: PINEDA, Omaira
; APPLICANT: YU, Guo-Liang
; APPLICANT: BROUN, Pierre E
; TITLE OF INVENTION: Yield-Related Polynucleotides and Polypeptides in Plants
; FILE REFERENCE: MBI0036-2 US
; CURRENT APPLICATION NUMBER: US/10/225,066A
; CURRENT FILING DATE: 2002-08-09
; PRIOR APPLICATION NUMBER: 09/837,444
; PRIOR FILING DATE: 2001-04-18
; PRIOR APPLICATION NUMBER: 60/310,847
; PRIOR FILING DATE: 2001-08-09
; PRIOR APPLICATION NUMBER: 60/336,049
; PRIOR FILING DATE: 2001-12-05
; PRIOR APPLICATION NUMBER: 60/338,692
; PRIOR FILING DATE: 2001-12-11
; PRIOR APPLICATION NUMBER: 10/171,468
; PRIOR FILING DATE: 2002-06-14
; NUMBER OF SEQ ID NOS: 1122
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 922
; LENGTH: 521
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
; US-10-225-066A-922

Query Match      60.6%; Score 43; DB 12; Length 521;
Best Local Similarity 57.1%; Pred. No. 88;
Matches      8; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY      3 LXLPNIXTQPIPLL 16
      |||::|||::|||
Db      89 LFLPSWVTQPLPQL 102

RESULT 8
US-10-278-173-36
; Sequence 36, Application US/10278173
; Publication No. US20030061637A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Cai-Zhong
; APPLICANT: Broun, Pierre
; APPLICANT: Riechmann, Jose-Luis
; APPLICANT: Pineda, Omaira
; APPLICANT: Zhang, James
; APPLICANT: Yu, Guo-Liang
; APPLICANT: Pilgrim, Marsha
; APPLICANT: Keddle, James
; APPLICANT: Heard, Jacqueline
; APPLICANT: Reuber, Lynne
; APPLICANT: Ratcliffe, Oliver
; APPLICANT: Adam, Luc
; APPLICANT: Samaha, Raymond
; TITLE OF INVENTION: POLYNUCLEOTIDES FOR ROOT TRAIT ALTERATION
; FILE REFERENCE: MBI-009
; CURRENT APPLICATION NUMBER: US/10/278,173
; CURRENT FILING DATE: 2002-10-21
; PRIOR APPLICATION NUMBER: US/09/533,392
; PRIOR FILING DATE: 2000-03-22
; PRIOR APPLICATION NUMBER: 60/125,814
; PRIOR FILING DATE: 1999-03-23
; NUMBER OF SEQ ID NOS: 177
; SOFTWARE: PatentIn Ver. 2.1

; GENERAL INFORMATION:
; APPLICANT: Mendel Biotechnology, Inc.
; APPLICANT: RATCLIFFE, Oliver
; APPLICANT: RIECHMANN, Jose Luis
; APPLICANT: ADAM, Luc J
; APPLICANT: DUBELL, Arnold T
; APPLICANT: HEARD, Jacqueline E
; APPLICANT: PILGRIM, Marsha L
; APPLICANT: JIANG, Cai-Zhong
; APPLICANT: REUBER, T. Lynne
; APPLICANT: CREELMAN, Robert A
; APPLICANT: PINEDA, Omaira
; APPLICANT: YU, Guo-Liang
; APPLICANT: BROUN, Pierre E
; TITLE OF INVENTION: Yield-Related Polynucleotides and Polypeptides in Plants
; FILE REFERENCE: MBI0036-2 US
; CURRENT APPLICATION NUMBER: US/10/225,066A
; CURRENT FILING DATE: 2002-08-09
; PRIOR APPLICATION NUMBER: 09/837,444
; PRIOR FILING DATE: 2001-04-18
; PRIOR APPLICATION NUMBER: 60/310,847
; PRIOR FILING DATE: 2001-08-09
; PRIOR APPLICATION NUMBER: 60/336,049
; PRIOR FILING DATE: 2001-12-05
; PRIOR APPLICATION NUMBER: 60/338,692
; PRIOR FILING DATE: 2001-12-11
; PRIOR APPLICATION NUMBER: 10/171,468
; PRIOR FILING DATE: 2002-06-14
; NUMBER OF SEQ ID NOS: 1122
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 922
; LENGTH: 521
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
; US-10-225-066A-922

Query Match      60.6%; Score 43; DB 12; Length 521;
Best Local Similarity 57.1%; Pred. No. 88;
Matches      8; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY      3 LXLPNIXTQPIPLL 16
      |||::|||::|||
Db      89 LFLPSWVTQPLPQL 102

RESULT 7
US-10-225-066A-922
; Sequence 922, Application US/10225066A
; Publication No. US20030226173A1
```

```
; SEQ ID NO 36
; LENGTH: 521
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; OTHER INFORMATION: G176
; US-10-278-173-36

Query Match          60.6%; Score 43; DB 14; Length 521;
Best Local Similarity 57.1%; Pred. No. 88;
Matches 8; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 3 LXLPNIXTQPIPLL 16
Db 89 LFLPSMVTQPLPQL 102

RESULT 9
US-10-278-536-34
; Sequence 34, Application US/10278536
; Publication No. US20030131386A1
; GENERAL INFORMATION:
; APPLICANT: Samaha, Raymond
; APPLICANT: Heard, Jacqueline
; APPLICANT: Jiang, Cai-Zhong
; APPLICANT: Pineda, Omalra
; APPLICANT: Reuber, Lynne
; APPLICANT: Riechmann, Jose-Luis
; APPLICANT: Yu, Guo-Liang
; APPLICANT: Keddle, James
; APPLICANT: Ratcliffe, Oliver
; APPLICANT: Pilgrim, Marsha
; APPLICANT: Adam, Luc
; APPLICANT: Broun, Pierre
; TITLE OF INVENTION: STRESS-INDUCED POLYNUCLEOTIDES
; FILE REFERENCE: MBI-011
; CURRENT FILING DATE: 2002-10-22
; PRIOR APPLICATION NUMBER: US/10/278,536
; PRIOR FILING DATE: 1999-03-23
; NUMBER OF SEQ ID NOS: 238
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 34
; LENGTH: 521
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; OTHER INFORMATION: G176
; US-10-278-536-34

Query Match          60.6%; Score 43; DB 14; Length 521;
Best Local Similarity 57.1%; Pred. No. 88;
Matches 8; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 3 LXLPNIXTQPIPLL 16
Db 89 LFLPSMVTQPLPQL 102

RESULT 10
US-10-374-780A-2244
; Sequence 2244, Application US/10374780A
; Publication No. US2004001997A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, Bradley K
; APPLICANT: Riechmann, Jose Luis
; APPLICANT: Jiang, Cai-Zhong
; APPLICANT: Heard, Jacqueline E
; APPLICANT: Haake, Volker
; APPLICANT: Creelman, Robert A
; APPLICANT: Ratcliffe, Oliver
; APPLICANT: Adam, Luc J
; APPLICANT: Reuber, T. Lynne

; APPLICANT: Keddle, James
; APPLICANT: Broun, Pierre B
; APPLICANT: Pilgrim, Marsha L
; APPLICANT: Dubell III, Arnold T
; APPLICANT: Pineda, Omalra
; APPLICANT: Yu, Guo-Liang
; TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES IN PLANTS
; FILE REFERENCE: MBI-0047 CIP
; CURRENT APPLICATION NUMBER: US/10/374,780A
; CURRENT FILING DATE: 2003-02-25
; PRIOR APPLICATION NUMBER: 09/837,944
; PRIOR FILING DATE: 2001-04-18
; PRIOR APPLICATION NUMBER: 60/310,847
; PRIOR FILING DATE: 2001-08-09
; PRIOR APPLICATION NUMBER: 09/934,455
; PRIOR FILING DATE: 2001-08-22
; PRIOR APPLICATION NUMBER: 60/336,049
; PRIOR FILING DATE: 2001-11-19
; PRIOR APPLICATION NUMBER: 60/338,692
; PRIOR FILING DATE: 2001-12-11
; PRIOR APPLICATION NUMBER: 10/171,468
; PRIOR FILING DATE: 2002-06-14
; PRIOR APPLICATION NUMBER: 10/225,066
; PRIOR FILING DATE: 2002-08-09
; PRIOR APPLICATION NUMBER: 10/225,067
; PRIOR FILING DATE: 2002-08-09
; PRIOR APPLICATION NUMBER: 10/225,068
; PRIOR FILING DATE: 2002-08-09
; NUMBER OF SEQ ID NOS: 2906
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 2244
; LENGTH: 521
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; OTHER INFORMATION: G176
; US-10-374-780A-2244

Query Match          60.6%; Score 43; DB 15; Length 521;
Best Local Similarity 57.1%; Pred. No. 88;
Matches 8; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 3 LXLPNIXTQPIPLL 16
Db 89 LFLPSMVTQPLPQL 102

RESULT 11
US-10-424-599-182136
; Sequence 182136, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 182136
; LENGTH: 79
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(79)
; OTHER INFORMATION: unsure at all xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MBT3847_135482C.1.pap
; US-10-424-599-182136
```

Query Match 57.7%; Score 41; DB 12; Length 79;  
Best Local Similarity 63.6%; Pred. No. 25;  
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 6 PNIXTQPIPL 16  
|||:|:|:|  
Db 61 PITSQPL 71

RESULT 12  
US-10-156-761-9332  
; Sequence 9332, Application US/10156761  
; Publication No. US20030119018A1  
; GENERAL INFORMATION:  
; APPLICANT: OMURA, SATOSHI  
; APPLICANT: IKEDA, HARUO  
; APPLICANT: ISHIKAWA, JUN  
; APPLICANT: HORIKAWA, HIROSHI  
; APPLICANT: SHIBA, TADAYOSHI  
; APPLICANT: SAKAKI, YOSHIYUKI  
; APPLICANT: HATTORI, MASAHIRO  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-262  
; CURRENT APPLICATION NUMBER: US/10/156,761  
; CURRENT FILING DATE: 2002-05-29  
; PRIOR APPLICATION NUMBER: JP 2001-204089  
; PRIOR FILING DATE: 2001-05-30  
; PRIOR APPLICATION NUMBER: JP 2001-272697  
; PRIOR FILING DATE: 2001-08-02  
; NUMBER OF SEQ ID NOS: 15109  
; SEQ ID NO 9332  
; LENGTH: 499  
; TYPE: PRT  
; ORGANISM: Streptomyces avermitilis  
US-10-156-761-9332

Query Match 57.7%; Score 41; DB 14; Length 499;  
Best Local Similarity 43.8%; Pred. No. 1.8e+02;  
Matches 7; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

QY 1 IALXLPNIXTQPIPL 16  
|||:|:|:|  
Db 185 VAIGVPVLTQPSVL 200

RESULT 13  
US-10-369-493-2154  
; Sequence 2154, Application US/10369493  
; Publication No. US20030233675A1  
; GENERAL INFORMATION:  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Goldman, Barry S.  
; APPLICANT: Chen, Xianfeng  
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES  
; FILE REFERENCE: 38-10(52052)B  
; CURRENT APPLICATION NUMBER: US/10/369,493  
; CURRENT FILING DATE: 2003-02-28  
; PRIOR APPLICATION NUMBER: US 60/360,039  
; PRIOR FILING DATE: 2002-02-21  
; NUMBER OF SEQ ID NOS: 47374  
; SEQ ID NO 2154  
; LENGTH: 1089  
; TYPE: PRT  
; ORGANISM: Schizosaccharomyces pombe  
; FEATURE:  
; NAME/KEY: unsure  
; LOCATION: (1)..(1089)  
; OTHER INFORMATION: unsure at all Xaa locations  
US-10-369-493-2154

Query Match 57.7%; Score 41; DB 15; Length 1089;  
Best Local Similarity 66.7%; Pred. No. 4e+02;  
Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 5 LPNIXTQPIPL 16  
|||:|:|:|  
Db 498 LPSIGKTPPL 509

RESULT 14  
US-09-801-368-368  
; Sequence 368, Application US/09801368  
; Patent No. US20020128250A1  
; GENERAL INFORMATION:  
; APPLICANT: Busby, Robert  
; APPLICANT: Cali, Brian  
; APPLICANT: Hecht, Peter  
; APPLICANT: Holtzman, Doug  
; APPLICANT: Madden, Kevin  
; APPLICANT: Maxon, Mary  
; APPLICANT: Milne, Todd  
; APPLICANT: No. US20020128250A1man, Thea  
; APPLICANT: Royer, John  
; APPLICANT: Salama, Sofie  
; APPLICANT: Sherman, Amir  
; APPLICANT: Silva, Jeff  
; APPLICANT: Summers, Eric  
; TITLE OF INVENTION: Methods for Improving Secondary Metabolite Production in Fungi  
; FILE REFERENCE: 109272.147  
; CURRENT APPLICATION NUMBER: US/09/801,368  
; CURRENT FILING DATE: 2001-03-07  
; PRIOR APPLICATION NUMBER: US 09/487,558  
; PRIOR FILING DATE: 2000-01-19  
; PRIOR APPLICATION NUMBER: US 60/160,587  
; PRIOR FILING DATE: 1999-10-20  
; NUMBER OF SEQ ID NOS: 440  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 368  
; LENGTH: 1579  
; TYPE: PRT  
; ORGANISM: Saccharomyces cerevisiae  
US-09-801-368-368

Query Match 57.7%; Score 41; DB 9; Length 1579;  
Best Local Similarity 54.5%; Pred. No. 5.9e+02;  
Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 5 LPNIXTQPIPL 15  
|||:|:|:|  
Db 132 LPKLTQPVFV 142

RESULT 15  
US-10-369-493-2000  
; Sequence 2000, Application US/10369493  
; Publication No. US20030233675A1  
; GENERAL INFORMATION:  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Goldman, Barry S.  
; APPLICANT: Chen, Xianfeng  
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES  
; FILE REFERENCE: 38-10(52052)B  
; CURRENT APPLICATION NUMBER: US/10/369,493  
; CURRENT FILING DATE: 2003-02-28  
; PRIOR APPLICATION NUMBER: US 60/360,039  
; PRIOR FILING DATE: 2002-02-21  
; NUMBER OF SEQ ID NOS: 47374  
; SEQ ID NO 2000  
; LENGTH: 1579

```

; TYPE: PRT
; ORGANISM: Saccharomyces cerevisiae
US-10-369-493-2000

Query Match          57.7%   Score 41;   DB 15;   Length 1579;
Best Local Similarity 54.5%   Pred. No. 5.9e+02;
Matches 6;   Conservative 3;   Mismatches 2;   Indels 0;   Gaps 0;

QY          5 LPNIXTQPIEL 15
           || : ||| :|
Db         132 LPKLSIQPVFV 142

```

Search completed: May 26, 2004, 19:19:03  
Job time : 2.33495 secs

Result No.	Score	Query		DB	ID	Description
		Match	Length			
1	44	62.0	317	2	T31897	hypothetical prote
2	43	60.6	309	2	T48481	SPR1-like protein
3	43	60.6	522	2	T31893	hypothetical prote
4	43	60.6	1256	2	AB2042	hypothetical prote
5	42	59.2	218	2	S29712	segmentation prote
6	42	59.2	432	2	BE2577	seryl-tRNA synthet
7	41	57.7	124	2	T26799	hypothetical prote
8	41	57.7	1072	2	T37742	serine threonine-p
9	41	57.7	1579	2	S59801	protein kinase SSK
10	41	57.7	3157	2	R70969	probable PPE prote
11	40	56.3	258	2	S35276	probable export pr
12	40	56.3	373	2	A56392	beta-galactoside a
13	40	56.3	429	2	I64048	serine-tRNA ligase
14	39	54.9	421	1	S14742	acid phosphatase (
15	39	54.9	423	1	S06167	acid phosphatase (
16	39	54.9	423	2	A33395	asparic aminopept
17	39	54.9	526	2	T48467	penicillin acylase
18	38	53.5	329	2	AG1130	hypothetical prote
19	38	53.5	333	2	AF2394	interleukin-1 rece
20	38	53.5	416	2	S33473	hypothetical prote
21	38	53.5	435	2	T46443	stai+ protein - fi
22	38	53.5	453	2	A43765	long-chain-fatty-a
23	38	53.5	556	2	G81798	long-chain-fatty-a
24	38	53.5	556	2	H81068	long-chain-fatty-a
25	38	53.5	562	2	AB0253	DNA topoisomerase
26	38	53.5	651	2	G64068	probable nucleic a
27	38	53.5	706	2	JQ2210	floculation suppl
28	38	53.5	766	1	S61694	diacylglycerol kin
29	38	53.5	750	2	T28793	

Best Local Similarity 57.1%; Pred. No. 3.9;  
Matches 8; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 3 LXLPLNIXTQPIPL 16  
DB 89 LFLPSMVTQPLPOL 102  
||||:|||||

RESULT 3  
T31893  
hypothetical protein C03A7.11 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 17-Mar-2000  
A:Accession: T31893  
R:Greco, T.; Bradshaw, H.; Elliott, G.  
submitted to the EMBL Data Library, July 1997  
A:Description: The sequence of C. elegans cosmid C03A7.  
A:Reference number: Z21096  
A:Accession: T31893  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-522 <GRE>  
A:Cross-references: EMBL:AF016451; PIDN:AB86005.1; GSPDB:GN00023; CESP:C03A7.11  
A:Experimental source: strain Bristol N2; clone C03A7  
C:Genetics:  
A:Gene: CESP:C03A7.11  
A:Map position: 5  
A:Introns: 53/3; 117/1; 184/1; 357/1; 399/1  
C:Superfamily: glucuronosyltransferase

Query Match 60.6%; Score 43; DB 2; Length 522;  
Best Local Similarity 46.7%; Pred. No. 7;  
Matches 7; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 1 LALXLPNIXTQPIPL 15  
DB 171 VTIGLPNIASQVPL 185  
: : ||||| : |||

RESULT 4  
AB2042  
hypothetical protein all1888 [imported] - Nostoc sp. (strain PCC 7120)  
C:Species: Nostoc sp. PCC 7120  
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120  
C:Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 09-Dec-2002  
A:Accession: AB2042  
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi  
Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.  
DNA Res. 8, 205-213, 2001  
A:Title: Complete Genomic  
A:Reference number: AB1807; MUID:21595285; PMID:11759840  
A:Accession: AB2042  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-1256 <KUR>  
A:Cross-references: GB:BA000019; PIDN:BAB73587.1; PID:gl7130978; GSPDB:GN00179  
A:Experimental source: strain PCC 7120  
C:Genetics:  
A:Gene: all1888

Query Match 60.6%; Score 43; DB 2; Length 1256;  
Best Local Similarity 66.7%; Pred. No. 19;  
Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 5 LPNIXTQPIPL 16  
DB 333 LDNVISQPIPL 344  
| : : |||||

RESULT 5  
S29712  
segmentation protein hairy - red flour beetle  
C:Species: Tribolium castaneum (red flour beetle)

C:Date: 19-Mar-1997 #sequence\_revision 01-Aug-1997 #text\_change 17-Mar-1999  
C:Accession: S29712  
R:Sommer, R.J.; Tautz, D.  
Nature 361, 448-450, 1993  
A:Title: Involvement of an orthologue of the Drosophila pair-rule gene hairy in segment  
A:Reference number: S29712; MUID:93156810; PMID:8429884  
A:Accession: S29712  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-218 <SOM>

Query Match 59.2%; Score 42; DB 2; Length 218;  
Best Local Similarity 56.2%; Pred. No. 4;  
Matches 9; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY 1 LALXLPNIXTQPIPL 16  
DB 148 IALVLTQGNASPIPL 163  
||||| : |||

RESULT 6  
E82577  
seryl-tRNA synthetase XF2286 [imported] - Xylella fastidiosa (strain 9a5c)  
C:Species: Xylella fastidiosa  
C:Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 02-Sep-2000  
A:Accession: E82577  
R:anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequen  
Nature 406, 151-157, 2000  
A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.  
A:Reference number: A82515; MUID:20365717; PMID:10910347  
A:Note: for a complete list of authors see reference number A59328 below  
A:Accession: E82577  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-432 <SIM>  
A:Cross-references: GB:AE004040; GB:AE003849; NID:g9107437; PIDN:AAF85085.1; GSPDB:GN001  
A:Experimental source: strain 9a5c  
R:Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; A  
Briones, M.R.S.; Bueno, M.R.P.; Camargo, L.E.A.; Canargo, D.M.; Carrier, H  
as-Neto, E.; Docena, C.; El-Dorri, H.; Facincani, A.P.; Ferreira, A.J.S.  
submitted to GenBank, June 2000  
A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohm  
J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; Laigr  
Chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, E  
A:Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.;  
F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A  
Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasak  
A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveir  
M.; Tshako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z  
A:Reference number: A59328  
A:Contents: annotation  
C:Genetics:  
A:Gene: XF2286  
C:Superfamily: serine-tRNA ligase

Query Match 59.2%; Score 42; DB 2; Length 432;  
Best Local Similarity 46.7%; Pred. No. 8.6;  
Matches 7; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 1 LALXLPNIXTQPIPL 15  
DB 106 VALGIFNLQQDVPL 120  
: || : ||| : |||

RESULT 7  
T26799  
hypothetical protein Y40H7A.11 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
A:Accession: T26799  
R:McMurray, A.  
submitted to the EMBL Data Library, November 1998  
A:Reference number: Z20268



A:Accession: T26799  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-124 <WIL>  
 A:Cross-references: EMBL:AL033510; PIDN:CAA22069.1; CESP:Y40H7A.11  
 A:Experimental source: clone Y40H7A  
 C:Genetics:  
 A:Gene: CESP:Y40H7A.11  
 A:Introns: 67/3

Query Match 57.7%; Score 41; DB 2; Length 124;  
 Best Local Similarity 37.5%; Pred. No. 3.2;  
 Matches 6; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

QY 1 IALXLPNIXTOPIPL 16  
 : : : : :  
 Db 17 VAIGMKVNRQAPVL 32

RESULT 8  
 T37742  
 serine threonine-protein kinase - fission yeast (Schizosaccharomyces pombe)  
 C:Species: Schizosaccharomyces pombe  
 C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 17-Mar-2000  
 C:Accession: T37742  
 R:Rieger, M.; Wood, V.; Rajandream, M.A.; Batrell, B.G.  
 submitted to the EMBL Data Library, January 1999  
 A:Reference number: Z21743  
 A:Accession: T37742  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-1072 <RIE>  
 A:Cross-references: EMBL:AL035248; PIDN:CAA22846.1; GSPDB:GN00066; SPDB:SPAC167.01  
 A:Experimental source: strain 972h; cosmid c167  
 C:Genetics:  
 A:Gene: spac167.01; SPDB:SPAC167.01  
 A:Map position: 1  
 A:Introns: 1015/3  
 C:Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase homolo

Query Match 57.7%; Score 41; DB 2; Length 1072;  
 Best Local Similarity 66.7%; Pred. No. 36;  
 Matches 9; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 5 LPNIXTOPIPL 16  
 : : : : :  
 Db 498 LPSIGKKPIPL 509

RESULT 9  
 S59801  
 protein kinase SSK2 (EC 2.7.1.1) - yeast (Saccharomyces cerevisiae)  
 N:Alternate names: protein N3276; protein YNR031C  
 C:Species: Saccharomyces cerevisiae  
 C:Date: 13-Jan-1996 #sequence\_revision 23-Feb-1996 #text\_change 24-Sep-1999  
 C:Accession: S59801; S63362  
 R:Maeda, T.; Takekawa, M.; Saito, H.  
 Science 269, 554-558, 1995  
 A:Title: Activation of yeast PBS2 MAPKK by MAPKKs or by binding of an SH3-containing os  
 A:Reference number: S59801; MUID:95350642; PMID:7624781  
 A:Accession: S59801  
 A:Status: nucleic acid sequence not shown  
 A:Molecule type: DNA  
 A:Residues: 1-1579 <MAE>  
 A:Cross-references: EMBL:L41927; NID:g940329; PIDN:AAC41665.1; PID:g940330  
 R:Pohl, T.M.  
 submitted to the Protein Sequence Database, April 1996  
 A:Reference number: S63346  
 A:Accession: S63362  
 A:Molecule type: DNA  
 A:Residues: 1-1579 <POH>  
 A:Cross-references: EMBL:271646; NID:g1302526; PIDN:CAA96311.1; PID:e239826; PID:g130252  
 A:Experimental source: strain S288C

C:Genetics:  
 A:Gene: SGD:SSK2  
 A:Cross-references: SGD:S0005314; MIPS:YNR031C  
 A:Map position: 14R  
 C:Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase homolo  
 C:Keywords: ATP; phosphotransferase; protein kinase  
 F:1284-1558/Domain: protein kinase homology <KIN>  
 F:1272-1280/Region: protein kinase ATP-binding motif

Query Match 57.7%; Score 41; DB 2; Length 1579;  
 Best Local Similarity 54.5%; Pred. No. 56;  
 Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 5 LPNIXTOPIPL 15  
 : : : : :  
 Db 132 LPKLSTQPPV 142

RESULT 10  
 B70969  
 Probable PPE protein - Mycobacterium tuberculosis (strain H37RV)  
 C:Species: Mycobacterium tuberculosis  
 C:Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 22-Oct-1999  
 C:Accession: B70969  
 R:Cole, S.T.; Broech, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.;  
 Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.;  
 Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.  
 Nature 393, 537-544, 1998  
 A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
 A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome  
 A:Reference number: A70500; MUID:98295987; PMID:9634230  
 A:Accession: B70969  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-3157 <COL>  
 A:Cross-references: GB:AL009198; GB:AL123456; NID:g3242262; PIDN:CAA15732.1; PID:e120226;  
 A:Experimental source: strain H37RV  
 C:Genetics:  
 A:Gene: PPE

Query Match 57.7%; Score 41; DB 2; Length 3157;  
 Best Local Similarity 53.8%; Pred. No. 1.2e+02;  
 Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 3 LXLNIXTOPIPL 15  
 : : : : :  
 Db 1789 VTPIITTSPIPL 1801

RESULT 11  
 S35276  
 Probable export protein mopC - Erwinia carotovora subsp. atroseptica  
 C:Species: Erwinia carotovora subsp. atroseptica  
 C:Date: 19-Mar-1997 #sequence\_revision 06-Jun-1997 #text\_change 29-Sep-1999  
 C:Accession: S35276  
 R:Mulholland, V.; Hinton, J.C.D.; Sidebotham, J.; Toth, I.K.; Hyman, L.J.; Perombelon, M.  
 Mol. Microbiol. 9, 343-356, 1993  
 A:Title: A pleiotropic reduced virulence (Rvi(-)) mutant of Erwinia carotovora subsp. at  
 gens.  
 A:Reference number: S35274; MUID:94018629; PMID:8412685  
 A:Accession: S35276  
 A:Molecule type: DNA  
 A:Residues: 1-258 <MUL>  
 A:Cross-references: EMBL:X72969; NID:g397346; PIDN:CAA51476.1; PID:g296055  
 A:Experimental source: strain SCRI 193  
 C:Genetics:  
 A:Gene: mopC  
 C:Function:  
 A:Description: may be involved in flagellar assembly; may be involved in export of flagel  
 C:Superfamily: flagellar biosynthetic protein flip  
 C:Keywords: transmembrane protein  
 F:58-74/Domain: transmembrane #status predicted <TM1>  
 F:101-117/Domain: transmembrane #status predicted <TM2>

F:202-218/Domain: transmembrane #status predicted <TM3>  
F:236-252/Domain: transmembrane #status predicted <TM4>

Query Match 56.3%; Score 40; DB 2; Length 258;  
Best Local Similarity 70.0%; Pred. No. 11;  
Matches 7; Conservative 1; Mismatches 0; Gaps 0;

QY 5 LPNIXTQPIIP 14  
||| |||||  
DB 36 LPGLIVTQPLP 45

## RESULT 12

A56392

beta-galactoside alpha1,2-fucosyltransferase I - rabbit

C:Species: Oryctolagus cuniculus (domestic rabbit)

C:Date: 19-Oct-1995 #sequence\_revision 19-Oct-1995 #text\_change 05-Nov-1999

C:Accession: A56392

R:Hitoshi, S.; Kusumoki, S.; Kanazawa, I.; Tsuji, S.

J. Biol. Chem. 270, 8844-8850, 1995

A:Title: Molecular cloning and expression of two types of rabbit beta-galactoside alpha1,2-fucosyltransferase I and II

A:Reference number: A56392; MUID:95238380; PMID:7721792

A:Accession: A56392

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-373 <HIT>

A:Cross-references: GB:X80226; NID:9854354; PIDN:CAA56513.1; PID:9854355

C:Keywords: transmembrane protein

Query Match 56.3%; Score 40; DB 2; Length 373;

Best Local Similarity 50.0%; Pred. No. 17;

Matches 7; Conservative 3; Mismatches 0; Indels 4; Gaps 0;

QY 1 IALXLPNIXTQPIIP 14

:|| ||| :|||

DB 37 LALSPLCLERQVPV 50

## RESULT 13

I64048

serine-tRNA ligase (EC 6.1.1.11) - Haemophilus influenzae (strain Rd KW20)

N:Alternate names: seryl-tRNA synthetase

C:Species: Haemophilus influenzae

C:Date: 18-Aug-1995 #sequence\_revision 18-Aug-1995 #text\_change 03-Jun-2002

C:Accession: I64048

R:Feischmann, R.B.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage, A.

; Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.F.; Glodek, A.; Kelley, J.M.; Weidman, J.

; D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann, J.L.; Geoghegan, N.S.M.

Science 269, 496-512, 1995

A:Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter, H.

A:Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.

A:Reference number: A54000; MUID:95350630; PMID:7542800

A:Accession: I64048

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-429 <TIGR>

A:Cross-references: GB:U32696; GB:L42023; NID:91573057; PIDN:AAC21787.1; PID:91573062; T

C:Genetics:

A:Gene: serS

C:Function:

A:Description: charges tRNA(Ser) with serine

A:Pathway: protein biosynthesis

C:Superfamily: serine-tRNA ligase

C:Keywords: aminoacyl-tRNA synthetase; ATP, ligase; protein biosynthesis

Query Match 56.3%; Score 40; DB 2; Length 429;

Best Local Similarity 46.7%; Pred. No. 20;

Matches 7; Conservative 3; Mismatches 5; Indels 5; Gaps 0;

QY 1 IALXLPNIXTQPIIP 15

:||| :|||

DB 100 IALSLPCLERQVPV 114

## RESULT 14

S14742

acid phosphatase (EC 3.1.3.2) precursor - mouse (fragment)

C:Species: Mus musculus (house mouse)

C:Date: 21-Nov-1993 #sequence\_revision 10-Nov-1995 #text\_change 31-Mar-2000

C:Accession: S14742

R:Geier, C.; von Figura, K.; Pohlmann, R.

Biol. Chem. Hoppe-Seyler 372, 301-304, 1991

A:Title: Molecular cloning of the mouse lysosomal acid phosphatase.

A:Reference number: S14742; MUID:91282986; PMID:2059337

A:Accession: S14742

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-421 <GEI>

A:Cross-references: EMBL:X57199; NID:952870; PIDN:CAA40485.1; PID:952871

C:Superfamily: mammalian acid phosphatase

C:Keywords: lysosome; phosphoric monoester hydrolase

Query Match 54.9%; Score 39; DB 2; Length 421;

Best Local Similarity 70.0%; Pred. No. 29;

Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 6 PNIXTQPIPL 15

|||||

DB 130 PNISWQPIPV 139

## RESULT 15

S06167

acid phosphatase (EC 3.1.3.2) ACP2 precursor [validated] - human

N:Alternate names: acid phosphatase, lysosomal

C:Species: Homo sapiens (man)

C:Date: 04-Dec-1992 #sequence\_revision 01-Dec-1995 #text\_change 08-Dec-2000

C:Accession: S06167; S05525; S01155

R:von Figura, K.

submitted to the EMBL Data Library, June 1989

A:Reference number: S06167

A:Accession: S06167

A:Molecule type: DNA

A:Residues: 1-423 <VON>

A:Cross-references: EMBL:X15525; NID:934239; PIDN:CAA33542.1; PID:91199524

R:Geier, C.; von Figura, K.; Pohlmann, R.

Eur. J. Biochem. 183, 611-616, 1989

A:Title: Structure of the human lysosomal acid phosphatase gene.

A:Reference number: S05525; MUID:89377828; PMID:2776754

A:Accession: S05525

A:Molecule type: DNA

A:Residues: 1-29 <GEI>

R:Pohlmann, R.; Krentler, C.; Schmidt, B.; Schroeder, W.; Lorkowski, G.; Culley, J.; Mer

EMBO J. 7, 2343-2350, 1988

A:Title: Human lysosomal acid phosphatase: cloning, expression and chromosomal assignment

A:Reference number: S01155; MUID:89052645; PMID:3191910

A:Accession: S01155

A:Molecule type: mRNA

A:Residues: 1-423 <POH>

A:Cross-references: EMBL:X12549; NID:934262; PIDN:CAA31064.1; PID:934263

A:Note: part of this sequence, including the amino end of the mature protein, was confir

C:Genetics:

A:Gene: GDB:ACP2

A:Cross-references: GDB:118963; OMIM:171650

A:Map position: 11p11.2-11p11.1

A:Introns: 38/3; 70/3; 99/3; 150/3; 183/3; 213/3; 258/1; 285/3; 321/2; 380/1

C:Function:

A:Description: catalyzes the hydrolysis of a wide range of phosphate esters

C:Superfamily: mammalian acid phosphatase

C:Keywords: glycoprotein, phosphohistidine, phosphoprotein; phosphoric monoester hydroly

F:1-30/Domain: signal sequence #status predicted <SIG>

F:31-423/Product: acid phosphatase ACP2 #status experimental <MAT>

F:41/Active site: Arg #status predicted

F:42/Active site: His (phosphohistidine intermediate) #status predicted

F:92,133,167,177,191,267,322,331/Binding site: carbohydrate (Asn) (covalent) #status pre

F:159-370,212-310,345-349/Disulfide bonds: #status predicted

Query Match 54.9%; Score 39; DB 1; Length 423;  
 Best Local Similarity 70.0%; Pred. No. 29;  
 Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 6 PNIXTQPIPL 15  
 ||| ||||:  
 Db 132 PNISWQPIPV 141

Search completed: May 26, 2004, 18:48:02  
 Job time : 1.44015 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 26, 2004, 17:20:49 ; Search time 0.256348 Seconds  
(without alignments)  
3249.964 Million cell updates/sec

Title: US-09-302-812-14

Perfect score: 71

Sequence: 1 IALXLPNIXTQIPILL 16

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_42.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	43	60.6	309	WR26 ARATH	Q9c5t3 arabidopsis
2	42	59.2	426	SYS_XANAC	Q8ply2 xanthomonas
3	42	59.2	426	SYS_XYLEFA	Q9pb58 xylella fas
4	42	59.2	426	SYS_XILET	Q87bx8 xylella fas
5	41	57.7	1579	SSK2 YEAST	P53599 saccharomyc
6	40	56.3	258	FLIP ERWCA	P34200 erwinnia car
7	40	56.3	373	FUT1 RABIT	Q10979 oryctolagus
8	40	56.3	429	SYS_HAETN	P43833 haemophilus
9	39	54.9	423	PPAL_HUMAN	P11117 homo sapien
10	39	54.9	423	PPAL_MOUSE	P24638 mus musculus
11	39	54.9	423	PPAL_RAT	P20611 rattus norv
12	38	53.5	416	ILIS RAT	P43303 rattus norv
13	38	53.5	453	STS1 SCHPO	P36209 schizosacch
14	38	53.5	562	LCFA_HAEIN	P46450 haemophilus
15	38	53.5	562	LCFA_YERPE	Q8ze89 yersinia pe
16	38	53.5	651	TOP3_HAEIN	P43704 haemophilus
17	38	53.5	766	SFL1 YEAST	P20134 saccharomyc
18	38	53.5	1087	POM1 SCHPO	Q09690 schizosacch
19	38	53.5	1239	CHS5 USMPA	O13394 ustilago ma
20	37	52.1	245	FLIP_ECOLI	P33133 escherichia
21	37	52.1	253	LEPA_VIBCH	P27717 vibrio chol
22	37	52.1	430	SYS_ECOLI	P09156 escherichia
23	37	52.1	430	SYS_SALTY	Q8xgn9 salmonella
24	37	52.1	561	LCFA_EC057	Q8xd6 escherichia
25	37	52.1	561	LCFA_EC011	P29212 escherichia
26	37	52.1	561	LCFA_SALTY	Q8xgn8 salmonella
27	37	52.1	639	V70K_PLRVN	P11622 potato leaf
28	37	52.1	651	PIAI_HUMAN	Q75925 homo sapien
29	37	52.1	651	PIAI_MOUSE	Q88907 mus musculus
30	36	50.7	313	LB36 ARATH	Q9fk33 arabidopsis
31	36	50.7	337	PLSX_VIBMA	Q9ra35 vibrio mari
32	36	50.7	420	SYS_MYCPN	P75107 mycoplasma
33	36	50.7	428	SYS_PASMU	P57836 pasteurilla

34	36	50.7	449	1	TIG_RALSO	Q8xyp8 ralstonia s
35	36	50.7	514	1	SADI SCHPO	Q09825 schizosacch
36	36	50.7	601	1	GLMS PYRAE	Q8zt20 p glucosami
37	36	50.7	611	1	ADAS_DICDI	Q96759 dictyosteli
38	36	50.7	662	1	LOX1_RABIT	P12530 oryctolagus
39	36	50.7	662	1	LOXP_RABIT	O19043 oryctolagus
40	36	50.7	687	1	YL18_CAEEL	Q11103 caenorhabdi
41	36	50.7	809	1	SYFB_RALSO	Q8xx24 ralstonia s
42	36	50.7	1381	1	Y3E7 YEAST	P34216 saccharomyc
43	35.5	50.0	501	1	DGRE_DROME	O44424 drosophila
44	35	49.3	193	1	DCD_BIFLO	Q8g478 bifidobacte
45	35	49.3	331	1	RDHD_HUMAN	Q8nbn7 homo sapien

#### ALIGNMENTS

#### RESULT 1

WR26 ARATH ID WR26 ARATH STANDARD; PRT; 309 AA.  
AC Q9C5T3; Q8GYK8; Q9LYQ5;  
DT 28-FEB-2003 (Rel. 41, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 15-MAR-2004 (Rel. 43, Last annotation update)  
DE Probable WRKY transcription factor 26 (WRKY DNA-binding protein 26)  
DE (SPF1-like protein).  
GN WRKY26 OR AT5G07100 OR MQJ9.27 OR T28J14\_40.  
OS Arabidopsis thaliana (Moune-ear cross).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;  
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.  
OX NCBI\_TaxID=3702;  
RN [1]  
RP SEQUENCE FROM N.A., AND INDUCTION.  
RX MEDLINE=21342502; PubMed=11449049;  
RA Yu D., Chen C., Chen Z.;  
RT "Evidence for an important role of WRKY DNA binding proteins in the regulation of NPR1 gene expression."  
RL Plant Cell 13:1527-1540(2001).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=cv. Columbia;  
RX MEDLINE=98344145; PubMed=9679202;  
RA Kaneko T., Kotani H., Nakamura Y., Sato S., Asamizu E., Miyajima N., Tabata S.;  
RT "Structural analysis of Arabidopsis thaliana chromosome 5. V. Sequence features of the regions of 1,381,565 bp covered by twenty one physically assigned pl and TAC clones."  
RL DNA Res. 5:131-145(1998).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=cv. Columbia;  
RX MEDLINE=21016721; PubMed=11130714;  
RA Miyajima N., Sasamoto S., Kimura T., Hosouchi T., Kawashima K., Kohara M., Matsumoto M., Matsuno A., Muraki A., Nakayama S., Watanabe N., Naruo K., Okumura S., Shino S., Takeuchi C., Wada T., Huang E., Spiegel L., Gnoj L., O'Shaughnessy A., Preston R., Habermann K., Murray J., Johnson D., Rohlfing T., Nelson J., Steneking T., Pepin K., Spieth J., Sekhon M., Armstrong J., Becker M., Belter E., Cordum H., Cordes M., Courtney L., Courtney W., Dante M., Du H., Edwards J., Fryman J., Haakensen B., Lamar E., Latreille P., Leonard S., Meyer R., Mulvaney E., Ozersky P., Riley A., Stromatt C., Wagner-McPherson C., Wollam A., Yeakum M., Bell M., Dedhia N., Parnell L., Shah R., Rodriguez M., Hoon See L., Vil D., Baker J., Kirchhoff K., Toth K., King L., Bahret A., Miller B., Marra M.A., Martensen R., McCombie W.R., Wilson R.K., Murphy G., Bancroft I., Volckaert G., Wambutt R., Huesterhoef A., Stiekema W., Pohl T., Entian K.-D., Terry N., Hartley N., Bent E., Johnson S., Langham S.-A., McCallagh B., Robben J., Grymonprez B., Zimmermann W., Ramsperger U., Wedler H., Balke K., Wedler E., Peters S., van Staveren M., Dirkse W., Moolijman P., Klein Lankhorst R.,

RA Weitzenecker T., Bothe G., Rose M., Hauf J., Berneiser S., Hempel S.,  
RA Feldpausch M., Lamberth S., Villarreal R., Gielen J., Ardiles W.,  
RA Bente O., Lemcke K., Kolesov G., Mayer K.F.X., Rudd S., Schoof H.,  
RA Schueller C., Zaccaria P., Mewes H.-W., Bevan M., Franz P.F.:  
RT "Sequence and analysis of chromosome 5 of the plant Arabidopsis  
RT thaliana";  
RL Nature 408:823-826 (2000).  
RN [4]  
RP SEQUENCE FROM N.A.  
RA Brover V., Troughan M., Alexandrov N., Lu Y.-P., Flavell R.,  
RA Feldmann K.A.;  
RT "Full-length cDNA from Arabidopsis thaliana";  
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.  
RN [5]  
RP SEQUENCE FROM N.A.  
RC STRAIN=cv. Columbia;  
RA Seki M., Iida K., Satou M., Sakurai T., Akiyama K., Ishida J.,  
RA Nakajima M., Enju A., Kamiya A., Narusaka M., Carninci P., Kawai J.,  
RA Hayashizaki Y., Shinozaki K.;  
RT "Arabidopsis thaliana full-length cDNA";  
RL Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.  
CC -!- FUNCTION: Transcription factor. Interacts specifically with the W  
CC box (5'-(T)TGAC(C/T)-3'), a frequently occurring elicitor-  
CC responsive cis-acting element (by similarity).  
CC -!- SUBCELLULAR LOCATION: Nuclear (Probable).  
CC -!- INDUCTION: By salicylic acid.  
CC -!- SIMILARITY: Belongs to the WRKY domain family.  
CC -!- SIMILARITY: Contains 2 WRKY domains.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
DR EMBL; AF224699; AAK28309.1; -;  
DR EMBL; AB010697; BAB11168.1; -;  
DR EMBL; AL163652; CAB87266.1; -;  
DR EMBL; AY084692; ABM61254.1; -;  
DR EMBL; AK117545; BAC42206.1; -;  
DR F1R; T48481; T48481.  
DR TRANSFAC; T05475; -;  
DR InterPro; IPR003657; WRKY.  
DR Pfam; PF03106; WRKY; 2.  
DR PROSITE; PS0811; WRKY; 2.  
KW Transcription regulation; DNA-binding; Nuclear protein; Repeat.  
FT DNA\_BIND 111 176 WRKY 1.  
FT DNA\_BIND 228 293 WRKY 2.  
FT CONFLICT 148 148 L -> P (IN REF. 5).  
FT CONFLICT 165 165 I -> F (IN REF. 4).  
FT CONFLICT 183 183 S -> P (IN REF. 1).  
SQ SEQUENCE 309 AA; 34910 MW; B83AF1B93F3909E CRC64;  
Query Match 60.6%; Score 43; DB 1; Length 309;  
Best Local Similarity 57.1%; Pred. No. 1.6;  
Matches 8; Conservative 3; Mismatches 3; Indels 0; Gaps 0;  
QY 3 LXLPLNXTQPIPL 16  
DB 89 LFLPSMVTQPLPQL 102  
RESULT 2  
ID SYS\_XANAC STANDARD; PRT; 426 AA.  
AC Q9PLX2;  
DT 10-OCT-2003 (Rel. 42, Created)  
DT 10-OCT-2003 (Rel. 42, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Seryl-tRNA synthetase (EC 6.1.1.11) (Serine--tRNA ligase) (SerRS).  
GN SERS OR XAC1653.

OS Xanthomonas axonopodis (pv. citri).  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;  
OC Xanthomonadaceae; Xanthomonas.  
OX NCBI\_TaxID=92829;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=306 / ATCC 13902 / XV 101;  
RX MEDLINE=22022145; PubMed=12024217;  
RA da Silva A.C.R., Ferro J.A., Reinach F.C., Farah C.S., Furlan L.R.,  
RA Quaggio L.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,  
RA Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.E.A.,  
RA Camarotte G., Camnanan F., Cardozo J., Chambergo F., Ciapina L.P.,  
RA Ciccarelli R.M.B., Coutinho L.L., Cursino-Santos J.R., El-Dorri H.,  
RA Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Ferro M.I.T.,  
RA Formighieri E.F., Franco M.C., Greggio C.C., Gruber A.,  
RA Katsuyama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,  
RA Locati E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,  
RA Martins E.C., Meidanis J., Menck C.F.M., Miyaki C.Y., Moon D.H.,  
RA Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,  
RA Pereira H.A., Rossi A., Sena J.A.D., Silva C., de Souza R.F.,  
RA Spinola L.A.F., Takita M.A., Tamura R.E., Teixeira E.C., Tezza R.I.D.,  
RA Trindade dos Santos M., Truffi D., Tsai S.M., White F.F.,  
RA Setubal J.C., Kitajima J.P.;  
RT "Comparison of the genomes of two Xanthomonas pathogens with differing  
RT host specificities";  
RL Nature 417:459-463 (2002).  
CC -!- CATALYTIC ACTIVITY: ATP + L-serine + tRNA(Ser) = AMP + diphosphate  
CC + L-seryl-tRNA(Ser).  
CC -!- SUBUNIT: Homodimer (by similarity).  
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.  
CC -!- SIMILARITY: Belongs to class-II aminoacyl-tRNA synthetase family.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
DR EMBL; AB011797; AAM36521.1; -;  
DR HAMAP; MF\_00176; 1.  
DR InterPro; IPR002314; tRNA-synt\_2b.  
DR InterPro; IPR002317; tRNA-synt\_ser.  
DR InterPro; IPR006195; tRNA\_ligase\_II.  
DR Pfam; PF02403; Seryl tRNA N; 1.  
DR Pfam; PF00587; tRNA-synt\_2b; 1.  
DR PRINTS; PR00981; TRNASYNTHSER.  
DR TIGRFAMs; TIGR00414; sers; 1.  
DR PROSITE; PS0862; AA tRNA LIGASE II; 1.  
KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;  
KW Complete proteome.  
SQ SEQUENCE 426 AA; 47118 MW; B72EBA3070F6F5A4 CRC64;  
Query Match 59.2%; Score 42; DB 1; Length 426;  
Best Local Similarity 53.3%; Pred. No. 3.6;  
Matches 8; Conservative 3; Mismatches 4; Indels 0; Gaps 0;  
QY 1 IALXLPNIXTOPIPL 15  
DB 100 IALGLNLPAGVPL 114  
RESULT 3  
ID SYS\_XYLEA STANDARD; PRT; 426 AA.  
AC Q9P558;  
DT 10-OCT-2003 (Rel. 42, Created)  
DT 10-OCT-2003 (Rel. 42, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Seryl-tRNA synthetase (EC 6.1.1.11) (Serine--tRNA ligase) (SerRS).  
GN SERS OR XF2286.  
OS Xylella fastidiosa.

OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;  
 OC Xanthomonadaceae; Xylella.

OX NCBI\_TaxID=2371;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=9abc;

RX MEDLINE=20365717; PubMed=10910347;

RA Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,

RA Alvaranga R., Alves L.M.C., Araya J.E., Baia G.S., Baptista C.S.,

RA Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Briones M.R.S.,

RA Bueno M.R.P., Camargo A.A., Carraro L.E.A., Carraro D.M., Carrier H.,

RA Colauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,

RA Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorry H.,

RA Faccinani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,

RA Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furlan L.R.,

RA Garnier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,

RA Ho P.L., Hoheisel J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,

RA Krieger J.E., Kuramae E.E., Laigret F., Lambais M.R., Leite L.C.C.,

RA Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,

RA Machado M.A., Madeira A.B.N., Madeira H.M.F., Marino C.L.,

RA Marques M.V., Martins A.L., Martins E.M.F., Matsukuma A.Y.,

RA Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,

RA Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,

RA Nhani A.J., Nobrega F.G., Nunes L.R., Oliveira M.A.,

RA de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,

RA Peixoto B.R., Pereira G.A.G., Pereira H.A. Jr., Pesquero J.B.,

RA Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,

RA de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,

RA da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,

RA da Silveira J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,

RA Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,

RA Zago M.A., Zatz M., Meidanis J., Setubal J.C.

RT "The genome sequence of the plant pathogen Xylella fastidiosa";

RL Nature 406:151-159(2000).

CC -1- CATALYTIC ACTIVITY: ATP + L-serine + tRNA(Ser) = AMP + diphosphate

CC + L-seryl-tRNA(Ser).

CC -1- SUBUNIT: Homodimer (By similarity).

CC -1- SUBCELLULAR LOCATION: Cytoplasmic.

CC -1- SIMILARITY: Belongs to class-II aminoacyl-tRNA synthetase family.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

CC EMBL; AE004040; AAF85085.1; ALT\_INIT.

CC PIR; E82577; E82577.

CC HSP; P34945; I5ER.

CC HAMAP; MF\_00176; -; 1.

CC InterPro; IPR002314; tRNA-synt\_2b.

CC InterPro; IPR002317; tRNA-synt\_ser.

CC InterPro; IPR006195; tRNA\_ligase\_II.

CC Pfam; PF02403; Seryl tRNA N; 1.

CC Pfam; PF00587; tRNA-synt\_2b; 1.

CC PRINTS; PR00981; TRNASYNTHSR.

CC TIGRFAMs; TIGR00414; sers; 1.

CC PROSITE; PS50862; AA\_TRNA\_LIGASE II; 1.

CC Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;

CC Complete proteome.

CC SEQUENCE 426 AA; 47425 MW; B21A58EC61592986 CRC64;

Query Match 59.2%; Score 42; DB 1; Length 426;

Best Local Similarity 46.7%; Pred. No. 3.6;

Matches 7; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

Qy 1 IALXLPNIXTQPIPL 15

Db 100 VALGIPLNPDQVPL 114

RESULT 4

IDS\_XYLFT

ID SYS\_XYLFT

AC Q87BX8;

DT 10-OCT-2003 (Rel. 42, Created)

DT 10-OCT-2003 (Rel. 42, Last sequence update)

DT 10-OCT-2003 (Rel. 42, Last annotation update)

DB Seryl-tRNA synthetase (EC 6.1.1.11) (Serine-tRNA ligase) (SerRS).

GN SRS OR PD1318.

OS Xylella fastidiosa (strain Temeculap / ATCC 700964).

OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;

OC Xanthomonadaceae; Xylella.

OX NCBI\_TaxID=183190;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=22421331; PubMed=12533478;

RA Van Sluys M.A., de Oliveira M.C., Monteiro-Vitorello C.B.,

RA Miyaki C.Y., Furlan L.R., Camargo L.E.A., da Silva A.C.R., Moon D.H.,

RA Takita M.A., Lemos E.G.M., Machado M.A., Ferro M.I.T., da Silva F.R.,

RA Goldman M.H.S., Goldman G.H., Lemos M.V.F., El-Dorry H., Tsai S.M.,

RA Carrier H., Carraro D.M., de Oliveira R.C., Nunes L.R., Siqueira W.J.,

RA Coutinho L.L., Kimura E.T., Ferro E.S., Harakava R., Kuramae E.E.,

RA Marino C.L., Gigliotti E., Abreu I.L., Alves L.M.C., do Amaral A.M.,

RA Baia G.S., Blanco S.R., Brito M.S., Cannavan F.S., Celestino A.V.,

RA da Cunha A.P., Fenille R.C., Ferro J.A., Formighieri E.F., Kishi L.T.,

RA Leoni S.G., Oliveira A.R., Rosa V.E. Jr., Sasaki F.T., Sena J.A.D.,

RA de Souza A.A., Truffi D., Teukumo F., Yanai G.M., Zaros L.G.,

RA Civerolo E.L., Simpson A.J.G., Almeida N.F. Jr., Setubal J.C.,

RA Kitajima J.P.

RT "Comparative analyses of the complete genome sequences of Pierce's

RT disease and citrus variegated chlorosis strains of Xylella

RT fastidiosa";

RL J. Bacteriol. 185:1018-1026(2003).

CC -1- CATALYTIC ACTIVITY: ATP + L-serine + tRNA(Ser) = AMP + diphosphate

CC + L-seryl-tRNA(Ser).

CC -1- SUBUNIT: Homodimer (By similarity).

CC -1- SUBCELLULAR LOCATION: Cytoplasmic.

CC -1- SIMILARITY: Belongs to class-II aminoacyl-tRNA synthetase family.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

CC EMBL; AE012558; AAC29167.1; -.

CC HAMAP; MF\_00176; -; 1.

CC InterPro; IPR002314; tRNA-synt\_2b.

CC InterPro; IPR002317; tRNA-synt\_ser.

CC InterPro; IPR006195; tRNA\_ligase\_II.

CC Pfam; PF02403; Seryl tRNA N; 1.

CC Pfam; PF00587; tRNA-synt\_2b; 1.

CC PRINTS; PR00981; TRNASYNTHSR.

CC PROSITE; PS50862; AA\_TRNA\_LIGASE II; 1.

CC Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;

CC Complete proteome.

CC SEQUENCE 426 AA; 47405 MW; E9C679A67B66B8E9 CRC64;

Query Match 59.2%; Score 42; DB 1; Length 426;

Best Local Similarity 46.7%; Pred. No. 3.6;

Matches 7; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

Qy 1 IALXLPNIXTQPIPL 15

Db 100 VALDIPLNPDQVPL 114

RESULT 5

SSK2\_YEAST

ID SSK2\_YEAST

PRT; 1579 AA.

AC P53599;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE MAP kinase kinase SSK2 (EC 2.7.1.-) (Suppressor of sensor  
 DE kinase 2).  
 GN SSK2 OR YNR031C OR N3276.  
 OS Saccharomyces cerevisiae (Baker's yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
 OX NCBI\_TaxID=4932;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=95350642; PubMed=7624781;  
 RA Maeda T., Takekawa M., Saito H.;  
 RT "Activation of yeast P52 MAPK by MAPKKs or by binding of an SH3-  
 RT containing osmosensor";  
 RL Science 269:554-558(1995).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Pohl T.M.;  
 RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.  
 CC -!- FUNCTION: Kinase involved in a signal transduction pathway that is  
 CC activated by changes in the osmolarity of the extracellular  
 CC environment. Activates the P52 MAP kinase kinase by  
 CC phosphorylation.  
 CC -!- SUBUNIT: Interacts with by SSK1.  
 CC -!- SIMILARITY: Belongs to the Ser/Thr family of protein kinases. MAP  
 CC kinase kinase subfamily.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 DR EMBL; L41927; AAC41665.1; -;  
 DR EMBL; Z71646; CAA96311.1; -;  
 DR PIR; S59801; S59801.  
 DR Germline; 143376; -;  
 DR SGD; S0005314; SSK2.  
 DR GO; GO:0005829; C:cytosol; IDA.  
 DR GO; GO:0030036; P:actin cytoskeleton organization and biogenesis; IDA.  
 DR InterPro; IPR000719; Prot kinase.  
 DR InterPro; IPR008271; Ser\_thr\_kinase.  
 DR InterPro; IPR002290; Ser\_thr\_kinase.  
 DR Pfam; PF00069; pkinase; 1.  
 DR ProDom; PD000001; Prot kinase; 1.  
 DR SMART; SM00220; S\_TKc; 1.  
 DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.  
 DR PROSITE; PS00108; PROTEIN KINASE ST; 1.  
 DR PROSITE; PS50011; PROTEIN KINASE DOM; 1.  
 KW Transferase; Serine/threonine-protein kinase; ATP-binding.  
 FT DOMAIN 1266 1558 PROTEIN KINASE.  
 FT NP\_BIND 1272 1280 ATP (BY SIMILARITY).  
 FT BINDING 1295 1295 ATP (BY SIMILARITY).  
 FT ACT\_SITE 1390 1390 BY SIMILARITY.  
 SQ SEQUENCE 1579 AA; 180526 MW; 91A07A195CB3772E CRC64;

Query Match 57.7%; Score 41; DB 1; Length 1579;  
 Best Local Similarity 54.5%; Pred. No. 24;  
 Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 5 LPNIXTQPIPL 15  
 || : ||| :|  
 Db 132 LPKLSQTPV 142

RESULT 6  
 FLIP\_ERWCA STANDARD; PRT; 258 AA.  
 ID FLIP\_ERWCA

AC P34200;  
 DT 01-FEB-1994 (Rel. 28, Created)  
 DT 01-FEB-1994 (Rel. 28, Last sequence update)  
 DT 15-JUL-1998 (Rel. 36, Last annotation update)  
 DE Flagellar biosynthetic protein flip (Flagellar biosynthetic protein  
 DE mopC).  
 GN FLIP OR MOPC.  
 OS Erwinia carotovora.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Pectobacterium.  
 OX NCBI\_TaxID=554;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX STRAIN=SCRI193;  
 RA MEDLINE=94018629; PubMed=8412685;  
 RA Mulholland V., Hinton J.C.D., Sidebotham J., Toth I.K.,  
 RA Hyman L.J., Perombelon M.C.M., Reeves P.J., Salmond G.P.C.;  
 RT "A pleiotropic reduced virulence (Rvi-) mutant of Erwinia carotovora  
 RT subspecies atroseptica is defective in flagella assembly proteins  
 RT that are conserved in plant and animal bacterial pathogens.";  
 RL Mol. Microbiol. 9:343-356(1993).  
 CC -!- FUNCTION: Plays a role in the flagellum-specific transport system  
 CC (By similarity).  
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).  
 CC -!- SIMILARITY: BELONGS TO THE FLIP/MOPC/SPAP FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 DR EMBL; X72969; CAA51476.1; -;  
 DR PIR; S35276; S35276.  
 DR InterPro; IPR005837; Flip.  
 DR InterPro; IPR005838; TypeIII\_P.  
 DR Pfam; PF00813; Flip; 1.  
 DR PRINTS; PR01302; TYPE3IMPPROT.  
 DR ProDom; PD002586; TypeIII\_P; 1.  
 DR TIGRFAM; TIGR01103; FLIP\_1.  
 DR PROSITE; PS01060; FLIP\_1; 1.  
 DR PROSITE; PS01061; FLIP\_2; 1.  
 KW Flagellum; Transmembrane.  
 FT TRANSMEM 20 40 POTENTIAL.  
 FT TRANSMEM 64 87 POTENTIAL.  
 FT TRANSMEM 98 118 POTENTIAL.  
 FT TRANSMEM 194 213 POTENTIAL.  
 FT TRANSMEM 220 242 POTENTIAL.  
 SQ SEQUENCE 258 AA; 28557 MW; BBA2BE838900A5C8 CRC64;

Query Match 56.3%; Score 40; DB 1; Length 258;  
 Best Local Similarity 70.0%; Pred. No. 4.8;  
 Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 5 LPNIXTQPIIP 14  
 || : ||| :|  
 Db 36 LFGIVTQPLP 45

RESULT 7  
 FUT1\_RABIT STANDARD; PRT; 373 AA.  
 ID FUT1\_RABIT

AC Q10979;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Galactoside 2-alpha-L-fucosyltransferase 1 (EC 2.4.1.69) (GDP-L-  
 DE fucose:beta-D-galactoside 2-alpha-L-fucosyltransferase 1)  
 DE (Alpha(1,2)F1) (Fucosyltransferase 1).  
 GN FUT1 OR RFT-1.  
 OS Oryctolagus cuniculus (Rabbit).

```

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95238380; PubMed=7721792;
RA Hitoehi S., Kusunoki S., Kanazawa I., Tsuji S.;
RT "Molecular cloning and expression of two types of rabbit beta-
RT galactoside alpha 1,2-fucosyltransferase."
RL J. Biol. Chem. 270:8844-8850(1995).
CC -|- CATALYTIC ACTIVITY: GDP-L-fucose + beta-D-galactosyl-R = GDP +
CC alpha-L-fucosyl-1,2-beta-D-galactosyl-R.
CC -|- PATHWAY: Glycosylation.
CC -|- SUBCELLULAR LOCATION: Type II membrane protein. Membrane-bound
CC form in trans cisternae of Golgi.
CC -|- TISSUE SPECIFICITY: Brain.
CC -|- MISCELLANEOUS: THERE ARE TWO GENES (FUT1 AND FUT2) WHICH ENCODE
CC GALACTOSIDE 2-L-FUCOSYLTRANSFERASE. THEY ARE EXPRESSED IN A
CC TISSUE-SPECIFIC MANNER WITH EXPRESSION RESTRICTED TO CELLS OF
CC MESODERMAL OR ENDODERMAL ORIGIN RESPECTIVELY.
CC -|- SIMILARITY: Belongs to the glycosyltransferase family 11.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; X80226; CAA56513.1; -.
DR PIR; A56392; A56392. Glyco trans 11.
DR InterPro; IPR002516; Glyco trans 11; 1.
DR Pfam; PF01531; Glyco trans 11; 1.
KW Transferase; Glycosyltransferase; Glycoprotein; Transmembrane;
KW Signal-anchor; Golgi stack.
FT DOMAIN 1 12 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 13 29 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT (POTENTIAL).
FT DOMAIN 30 373 LUMENAL, CATALYTIC (POTENTIAL).
FT CARBOHYD 66 66 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 301 301 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 327 327 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 373 AA; 42098 MW; 0A47A1786231525C CRC64;

Query Match 56.3%; Score 40; DB 1; Length 373;
Best Local Similarity 50.0%; Pred. No. 7.3;
Matches 7; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 1 IALXLPNIXTOPIIP 14
Db 37 LALSPLCLERQVPV 50

RESULT 8
SYS_HAEN
ID SYS_HAEN STANDARD; PRT; 429 AA.
AC P43833;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Seryl-tRNA synthetase (EC 6.1.1.11) (Serine--tRNA ligase) (SerRS).
GN SERS OR HI0110.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC Pasteurellaceae; Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=RD / KW20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,

```

```

RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghegan N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus influenzae
RT Rd."
RL Science 269:496-512(1995).
CC -|- CATALYTIC ACTIVITY: ATP + L-serine + tRNA(Ser) = AMP + diphosphate
CC + L-seryl-tRNA(Ser).
CC -|- SUBUNIT: Homodimer (By similarity).
CC -|- SUBCELLULAR LOCATION: Cytoplasmic.
CC -|- SIMILARITY: Belongs to class-II aminoacyl-tRNA synthetase family.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; U32696; AAC21787.1; -.
DR PIR; I64048; I64048.
DR HSSP; F34945; I5ER.
DR TIGR; HI0110; -.
DR HAMAP; MF 00176; -. 1.
DR InterPro; IPR002314; tRNA-synt 2b.
DR InterPro; IPR002317; tRNA-synt_ser.
DR InterPro; IPR006195; tRNA_ligase_II.
DR Pfam; PF02403; Seryl tRNA N; 1.
DR Pfam; PF00587; tRNA-synt 2b; 1.
DR PRINTS; PR00981; TRNASYNTHSER.
DR TIGRFAMs; TIGR00414; serS; 1.
DR PROSITE; PS50862; AA tRNA LIGASE II; 1.
KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
KW Complete proteome.
SQ SEQUENCE 429 AA; 47721 MW; 763967DFCD9838ED CRC64;

Query Match 56.3%; Score 40; DB 1; Length 429;
Best Local Similarity 46.7%; Pred. No. 8.5;
Matches 7; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 1 IALXLPNIXTOPIIP 15
Db 100 IALSIPNLPADEVPL 114

RESULT 9
PPAL HUMAN
ID PPAL HUMAN STANDARD; PRT; 423 AA.
AC P11117; Q9BTU7;
DT 01-JUL-1989 (Rel. 11, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Lysosomal acid phosphatase precursor (EC 3.1.3.2) (LAP).
GN ACP2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC TISSUE=Placenta;
RX MEDLINE=89052645; PubMed=3191910;
RA Pohlmann R., Krentler C., Schmidt B., Schroeder W., Lorkowski G.,
RA Culley J., Mersmann G., Geier C., Waheed A., Gottschalk S.,
RA Grzeschik K.H., Hasikik A., von Figura K.;
RT "Human lysosomal acid phosphatase: cloning, expression and
RT chromosomal assignment."

```



EMBO J. 7:2343-2350(1988).  
 [2]  
 RN SEQUENCE FROM N.A.  
 RP TISSUE=Leukocyte;  
 RC MEDLINE=99377828; PubMed=2776754;  
 RA Geier C., von Figura K., Pohlmann R.;  
 RT "Structure of the human lysosomal acid phosphatase gene";  
 RL Eur. J. Biochem. 183:611-616(1989).  
 [3]  
 RN SEQUENCE FROM N.A.  
 RP TISSUE=Kidney;  
 RC MEDLINE=22388257; PubMed=12477932;  
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Schenmen C.M., Schuler G.D.,  
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
 RA Hopkins R.F., Zeeberg B., Moore T., Max S.I., Wang J., Hsieh F.,  
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,  
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,  
 RA Bosak S.A., McGowan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahney J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,  
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickinson G.G.,  
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickinson G.G.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalhus D.E.,  
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length human  
 and mouse cDNA sequences";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 CC -1- CATALYTIC ACTIVITY: An orthophosphoric monoester + H(2)O = an  
 alcohol + phosphate.  
 CC -1- SUBCELLULAR LOCATION: Lysosomal.  
 CC -1- SIMILARITY: Belongs to the histidine acid phosphatase family.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 DR EMBL; X12548; CA331064.1; -;  
 DR EMBL; X15525; CA33542.1; -;  
 DR EMBL; X15526; CA33542.1; JOINED.  
 DR EMBL; X15527; CA33542.1; JOINED.  
 DR EMBL; X15528; CA33542.1; JOINED.  
 DR EMBL; X15529; CA33542.1; JOINED.  
 DR EMBL; X15530; CA33542.1; JOINED.  
 DR EMBL; X15531; CA33542.1; JOINED.  
 DR EMBL; X15532; CA33542.1; JOINED.  
 DR EMBL; X15533; CA33542.1; JOINED.  
 DR EMBL; X15534; CA33542.1; JOINED.  
 DR EMBL; X15535; CA33542.1; JOINED.  
 DR EMBL; BC003160; AA03160.1; -;  
 DR PIR; S06167; S06167.  
 DR HSP; P20646; IRPA.  
 DR Genew; HGNC:123; ACP2.  
 DR MIN; 171650; -;  
 DR GO; GO:0016021; C:integral to membrane; TAS.  
 DR GO; GO:0003993; F:acid phosphatase activity; TAS.  
 DR InterPro; IPR000560; HsAc phosphatase.  
 DR Pfam; PF00328; acid phosphat; 1.  
 DR PROSITE; PS00616; HIS-ACID PHOSPHAT 1; 1.  
 DR PROSITE; PS00778; HIS-ACID PHOSPHAT 2; 1.  
 KW Hydrolase; Signal; Glycoprotein; Lysosome.  
 FT SIGNAL 1 30  
 FT CHAIN 31 423  
 FT DISULFID 159 370  
 FT BY SIMILARITY.  
 FT DISULFID 212 310

FT DISULFID 345 349  
 FT ACT SITE 42 42  
 FT ACT SITE 84 84  
 FT ACT SITE 286 286  
 FT CARBOHYD 92 92  
 FT CARBOHYD 133 133  
 FT CARBOHYD 167 167  
 FT CARBOHYD 177 177  
 FT CARBOHYD 191 191  
 FT CARBOHYD 267 267  
 FT CARBOHYD 322 322  
 FT CARBOHYD 331 331  
 FT CONFLICT 29 29  
 SQ SEQUENCE 423 AA; 48316 MW; 1A2CFB40ED3724B CRC64;  
 Query Match 54.9%; Score 39; DB 1; Length 423;  
 Best Local Similarity 70.0%; Pred. No. 13;  
 Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
 QY 6 PNIXTQPIPL 15  
 DB 132 PNISWQPIPV 141  
 RESULT 10  
 ID PPAL\_MOUSE STANDARD; PRT; 423 AA.  
 AC P24638; Q8QUT5;  
 DT 01-MAR-1992 (Rel. 21, Created)  
 DT 10-OCT-2003 (Rel. 42, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Lysosomal acid phosphatase precursor (EC 3.1.3.2) (LAP).  
 GN ACP2.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Liver;  
 RC MEDLINE=22388257; PubMed=12477932;  
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Schenmen C.M., Schuler G.D.,  
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Moore T., Max S.I., Wang J., Hsieh F.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,  
 RA Bosak S.A., McGowan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahney J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,  
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickinson G.G.,  
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickinson G.G.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalhus D.E.,  
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length human  
 and mouse cDNA sequences";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 RN [2]  
 RP SEQUENCE OF 3-423 FROM N.A.  
 RC MEDLINE=91282986; PubMed=2059337;  
 RA Geier C., von Figura K., Pohlmann R.;  
 RT "Molecular cloning of the mouse lysosomal acid phosphatase";  
 RL Biol. Chem. Hoppe-Seyler 372:301-304(1991).  
 CC -1- CATALYTIC ACTIVITY: An orthophosphoric monoester + H(2)O = an  
 alcohol + phosphate.  
 CC -1- SUBCELLULAR LOCATION: Lysosomal.  
 CC -1- SIMILARITY: Belongs to the histidine acid phosphatase family.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

DR EMBL; BC023343; AAH23343.1; -  
 DR EMBL; X57199; CAA40485.1; -  
 DR F1R; S14742; S14742.  
 DR HSP; P15309; 2HPA.  
 DR MGI; 87882; Acp2.  
 DR InterPro; IPR000560; HisAc\_phsphatase.  
 DR Pfam; PF00328; acid\_phosphat\_1; 1.  
 DR PROSITE; PS00616; HIS\_ACID\_PHOSPHAT\_1; 1.  
 DR PROSITE; PS00778; HIS\_ACID\_PHOSPHAT\_2; 1.  
 KW Hydrolase; Signal; Glycoprotein; Lysosome.  
 FT SIGNAL 1 30  
 FT CHAIN 31 423 LYSOSOMAL ACID PHOSPHATASE.  
 FT DISULFID 159 370 BY SIMILARITY.  
 FT DISULFID 212 310 BY SIMILARITY.  
 FT DISULFID 345 349 BY SIMILARITY.  
 FT ACT SITE 42 42 NUCLEOPHILIC ACCEPTOR (BY SIMILARITY).  
 FT ACT SITE 84 84 BY SIMILARITY.  
 FT ACT SITE 286 286 PROTON DONOR (BY SIMILARITY).  
 FT CARBOHYD 92 92 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 133 133 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 167 167 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 177 177 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 191 191 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 197 197 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 267 267 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 322 322 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 331 331 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 374 374 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CONFLICT 131 131 N -> S (IN REF. 2).  
 SQ SEQUENCE 423 AA; 48508 MW; AFD7A5C90A4F2AF4 CRC64;

Query Match 54.9%; Score 39; DB 1; Length 423;  
 Best Local Similarity 70.0%; Pred. No. 13;  
 Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Oy 6 PNIXTOPIPL 15  
 ||| |||:  
 Db 132 PNISWQPIPV 141

RESULT 11  
 ID PPAL RAT STANDARD; PRT; 423 AA.  
 AC P20611; 1991 (Rel. 17, Created)  
 DT 01-FEB-1991 (Rel. 17, Last sequence update)  
 DT 01-FEB-1991 (Rel. 32, Last annotation update)  
 DE Lysosomal acid phosphatase precursor (EC 3.1.3.2) (LAP).  
 GN ACP2.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=89350910; PubMed=2764916;  
 RA Hmeno M., Fujita H., Noguchi Y., Kono A., Kato K.;  
 RT "Isolation and sequencing of a cDNA clone encoding acid phosphatase  
 in rat liver lysosomes."  
 RL Biochem. Biophys. Res. Commun. 162:1044-1053 (1989).  
 CC -1- CATALYTIC ACTIVITY: An orthophosphoric monoester + H(2)O = an  
 CC alcohol + phosphate.  
 CC -1- SUBCELLULAR LOCATION: Lysosomal.  
 CC -1- SIMILARITY: Belongs to the histidine acid phosphatase family.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

DR EMBL; M27893; AAA0744.1; -  
 DR F1R; A33395; A33395.  
 DR HSP; P20646; 1RPA.  
 DR InterPro; IPR000560; HisAc\_phsphatase.  
 DR Pfam; PF00328; acid\_phosphat\_1; 1.  
 DR PROSITE; PS00616; HIS\_ACID\_PHOSPHAT\_1; 1.  
 DR PROSITE; PS00778; HIS\_ACID\_PHOSPHAT\_2; 1.  
 KW Hydrolase; Signal; Glycoprotein; Lysosome.  
 FT SIGNAL 1 30  
 FT CHAIN 31 423 LYSOSOMAL ACID PHOSPHATASE.  
 FT DISULFID 159 370 BY SIMILARITY.  
 FT DISULFID 212 310 BY SIMILARITY.  
 FT DISULFID 345 349 BY SIMILARITY.  
 FT ACT SITE 42 42 NUCLEOPHILIC ACCEPTOR (BY SIMILARITY).  
 FT ACT SITE 84 84 BY SIMILARITY.  
 FT ACT SITE 286 286 PROTON DONOR (BY SIMILARITY).  
 FT CARBOHYD 92 92 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 133 133 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 167 167 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 177 177 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 191 191 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 197 197 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 267 267 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 322 322 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 331 331 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 374 374 N-LINKED (GLCNAC. .) (POTENTIAL).  
 SQ SEQUENCE 423 AA; 48319 MW; 6F4C4819CADD4496 CRC64;

Query Match 54.9%; Score 39; DB 1; Length 423;  
 Best Local Similarity 70.0%; Pred. No. 13;  
 Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Oy 6 PNIXTOPIPL 15  
 ||| |||:  
 Db 132 PNISWQPIPV 141

RESULT 12  
 ID ILIS RAT STANDARD; PRT; 416 AA.  
 AC P43303;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Interleukin-1 receptor, type II precursor (IL-1R-2).  
 GN IL1R2 OR IL1RB OR IL-1R2.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=95035882; PubMed=7524717;  
 RA Bristulf J., Gatti S., Malinowsky D., Bjork L., Sundgren A.K.,  
 RA Bartfai T.,  
 RT "Interleukin-1 stimulates the expression of type I and type II  
 interleukin-1 receptors in the rat insulinoma cell line Rins5F;  
 sequencing a rat type II interleukin-1 receptor cDNA."  
 RL Eur. Cytokine Netw. 5:319-330 (1994).  
 CC -1- FUNCTION: RECEPTOR FOR INTERLEUKIN-1 ALPHA (IL-1A), BETA (IL-1B),  
 CC AND INTERLEUKIN-1 RECEPTOR ANTAGONIST PROTEIN (IL-1RA).  
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.  
 CC -1- SIMILARITY: Belongs to the interleukin-1 receptor family.  
 CC -1- SIMILARITY: Contains 3 immunoglobulin-like C2-type domains.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement. See <http://www.isb-sib.ch/announcement/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch).

DR EMBL: Z22812; CAA80465.1; --  
 DR PIR: S33473; S33473.  
 DR InterPro: IPR007110; Ig-like.  
 DR InterPro: IPR003599; Ig.  
 DR InterPro: IPR004074; IIL\_receptorI/II.  
 DR InterPro: IPR004077; IIL\_receptorIIP.  
 DR Pfam: PF00047; Ig; 3.  
 DR PRINTS: PR01539; INTRLEUKINR2.  
 DR PRINTS: PR01536; INTRLEUKINR12F.  
 DR SMART: SMO0409; IG; 3.  
 DR PROSITE: PS00835; IG-LIKE; 3.  
 KW Immunoglobulin domain; Receptor; Glycoprotein; Transmembrane; Signal;  
 KW Repeat.  
 FT SIGNAL 1 13 POTENTIAL.  
 FT CHAIN 14 416 INTERLEUKIN-1 RECEPTOR, TYPE II.  
 FT DOMAIN 14 355 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 356 381 POTENTIAL.  
 FT DOMAIN 382 416 CYTOPLASMIC (POTENTIAL).  
 FT DOMAIN 29 136 IG-LIKE C2-TYPE 1.  
 FT DOMAIN 146 233 IG-LIKE C2-TYPE 2.  
 FT DOMAIN 249 357 IG-LIKE C2-TYPE 3.  
 FT DISULFID 64 120 BY SIMILARITY.  
 FT DISULFID 164 219 BY SIMILARITY.  
 FT DISULFID 270 338 BY SIMILARITY.  
 FT CARBOHYD 124 124 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 208 208 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 231 231 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 289 289 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 416 AA; 46353 MW; 50EFBA48881C5D4E CRC64;

Query Match 53.5%; Score 38; DB 1; Length 416;  
 Best Local Similarity 43.8%; Pred. No. 19;  
 Matches 7; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

Qy 1 IALXLPNIXTQIPILL 16  
 Db 236 IELRVKGITTEPIVI 251

## RESULT 13

ID STS1 SCHPO STANDARD; PRT; 453 AA.  
 AC P36209; O13891;  
 DT 01-JUN-1994 (Rel. 29, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Delta(24(24(1)))-sterol reductase (EC 1.3.1.71) (Sterol delta(24(28))-  
 DE reductase) (C-24(28) sterol reductase).  
 GN STS1 OR SPAC20G4.07C.  
 OS Schizosaccharomyces pombe (Fission yeast).  
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;  
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;  
 OC Schizosaccharomycetes.  
 OX NCBI\_TaxID=4896;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=92329994; PubMed=1320960;  
 RA Shimamura M., Goebel M., Yanagida M., Toda T.;  
 RT "Fission yeast sts1+ gene encodes a protein similar to the chicken  
 RT lamin B receptor and is implicated in pleiotropic drug-sensitivity,  
 RT divalent cation-sensitivity, and osmoregulation."  
 RL Mol. Biol. Cell 3:263-273(1992).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=972;  
 RX MEDLINE=21848401; PubMed=11859360;  
 RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,

RA Sgouras J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,  
 RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,  
 RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,  
 RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,  
 RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,  
 RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,  
 RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,  
 RA Oliver K., O'Neil S., Pearson D., Quail M.A., Rabinowitsch E.,  
 RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,  
 RA Skelton J., Simmonds M., Squares R., Stevens S., Warren T., Whitehead S.,  
 RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,  
 RA Woodward J., Volkart G., Aert R., Robben J., Grymonprez B.,  
 RA Welter J., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,  
 RA Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,  
 RA Borzym K., Langer I., Beck A., Lehtach H., Reinhardt R., Pohl T.M.,  
 RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,  
 RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,  
 RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,  
 RA Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,  
 RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,  
 RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,  
 RA Carrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,  
 RA Shipakowski G.V., Ussery D., Barrell B.G., Nurse P.;  
 RT "The genome sequence of Schizosaccharomyces pombe."  
 RN Nature 415:871-880(2002).  
 RP [3]  
 CH CHARACTERIZATION.  
 MD MEDLINE=94171077; PubMed=8125337;  
 LA Lai M.H., Bard M., Pierson C.A., Alexander J.F., Goebel M.,  
 CA Carter G.T., Kirsch D.R.;  
 RT "The identification of a gene family in the Saccharomyces cerevisiae  
 RT ergosterol biosynthesis pathway."  
 RN Gene 140:41-49(1994).  
 CC -1- CATALYTIC ACTIVITY: Ergosterol + NADP(+) = ergosta-  
 CC 5,7,22,24(24(1))-tetraen-3-beta-ol + NADPH.  
 CC -1- PATHWAY: Ergosterol biosynthesis.  
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Probable).  
 CC -1- SIMILARITY: Belongs to the ERG4/ERG24 family.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: X63549; CAA45113.1; --  
 DR EMBL: Z98600; CAB11256.1; --  
 DR PIR: A43765; A43765.  
 DR GenDB SPombe; SPAC20G4.07C; --  
 DR InterPro: IPR001171; ERG4 ERG24.  
 DR Pfam: PF01222; ERG4 ERG24; 1.  
 DR PROSITE: PS01017; STEROL\_REDUCT\_1; 1.  
 DR PROSITE: PS01018; STEROL\_REDUCT\_2; 1.  
 DR Sterol biosynthesis; Oxidoreductase; NADP; Transmembrane.  
 KW TRANSMEM 13 33 POTENTIAL.  
 FT TRANSMEM 77 97 POTENTIAL.  
 FT TRANSMEM 118 138 POTENTIAL.  
 FT TRANSMEM 153 173 POTENTIAL.  
 FT TRANSMEM 209 229 POTENTIAL.  
 FT TRANSMEM 278 298 POTENTIAL.  
 FT TRANSMEM 311 331 POTENTIAL.  
 FT TRANSMEM 399 419 POTENTIAL.  
 FT CONFLICT 412 412 C -> S (IN REF. 1).  
 SQ SEQUENCE 453 AA; 52545 MW; 4740B6E3BBD27CF CRC64;

Query Match 53.5%; Score 38; DB 1; Length 453;  
 Best Local Similarity 53.8%; Pred. No. 21;  
 Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

Qy 3 IALXLPNIXTQIPILL 15

Db 92 LTLPGVTRQSLPL 104

## RESULT 14

LCFA\_HAEIN STANDARD; PRT; 562 AA.  
 AC P46450;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Long-chain-fatty-acid-CoA ligase (EC 6.2.1.3) (Long-chain acyl-CoA synthetase).  
 GN FADD OR HI0390.1.  
 OS Haemophilus influenzae.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;  
 OC Pasteurellaceae; Haemophilus.  
 OX NCBI\_TaxID=727;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=rd / KW20 / ATCC 51907;  
 RX MEDLINE=95350630; PubMed=7542800;  
 RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F., Kierlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M., McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D., Scott J.D., Shirley R., Liu L.-T., Glodek A., Kelley J.M., Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D., Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C., Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M., Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O., Venter J.C.;  
 RT "Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.";  
 RL Science 269:496-512(1995).  
 RN [2]  
 RP IDENTIFICATION.  
 RA Koonin E.V., Rudd K.E.;  
 RL Submitted (SEP-1995) to Swiss-Prot.  
 CC -!- FUNCTION: Esterification, concomitant with transport, of exogenous long-chain fatty acids into metabolically active CoA thioesters for subsequent degradation or incorporation into phospholipids (By similarity).  
 CC -!- CATALYTIC ACTIVITY: ATP + a long-chain carboxylic acid + CoA = AMP + diphosphate + an acyl-CoA.  
 CC -!- COFACTOR: Magnesium (By similarity).  
 CC -!- SUBUNIT: Homodimer (Probable).  
 CC -!- SUBCELLULAR LOCATION: Partially membrane-associated (Potential).  
 CC -!- SIMILARITY: Belongs to the ATP-dependent AMP-binding enzyme family.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC EMBL; U32722; AAC22049.1; -  
 CC HSSP; P08659; 1LCI.  
 DR TIGR; HI0390.1; -  
 DR InterPro; IPR000873; AMP-bind.  
 DR Pfam; PF00501; AMP-binding; 1.  
 DR PROSITE; PS00455; AMP BINDING; 1.  
 DR Ligase; Fatty acid metabolism; Magnesium; Membrane; Complete proteome.  
 SQ SEQUENCE 562 AA; 63478 MW; 74BE7C8E9D711F12 CRC64;

Query Match 53.5%; Score 38; DB 1; Length 562;  
 Best Local Similarity 46.7%; Pred. No. 27;  
 Matches 7; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 1 IALXLPNIXTQPL 15

:|:|:|:|:|

Db 77 VALMMENILQYPAL 91

## RESULT 15

LCFA\_YERPE STANDARD; PRT; 562 AA.  
 AC Q8ZES9;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Long-chain-fatty-acid-CoA ligase (EC 6.2.1.3) (Long-chain acyl-CoA synthetase).  
 GN FADD OR YPO2074 OR Y2236.  
 OS Yersinia pestis.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Yersinia.  
 OX NCBI\_TaxID=632;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CO-92 / Biovar Orientalis;  
 RX MEDLINE=21470413; PubMed=11586360;  
 RA Parkhill J., Wren B.W., Thomson N.R., Titball R.W., Holden M.T.G., Prentice M.B., Sebahia M., James K.D., Churcher C., Mungall K.L., Baker S., Basham D., Bentley S.D., Brooks K., Cerdeno-Tarraga A.M., Chillingworth T., Cronin A., Davies R.M., Davis P., Dougan G., Felwell T., Hamlin N., Holroyd S., Jagels K., Karlyshev A.V., Leather S., Moule S., Oyston P.C.F., Quail M.A., Rutherford K., Simmonds M., Skelton J., Stevens K., Whitehead S., Barrell B.G.;  
 RT "Genome sequence of Yersinia pestis, the causative agent of plague.";  
 RL Nature 413:523-527(2001).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-KIMS / Biovar Mediaevalis;  
 RX MEDLINE=22137863; PubMed=12142430;  
 RA Deng W., Burland V., Plunkett G. III, Boutin A., Mayhew G.F., Liss P., Perna N.T., Rose D.J., Mau B., Zhou S., Schwartz D.C., Fetherston J.D., Lindler L.E., Brubaker R.R., Plano G.V., Straley S.C., McDonough K.A., Nilles M.L., Matson J.S., Blattner F.R., Perry R.D.;  
 RT "Genome sequence of Yersinia pestis KIM.";  
 RL J. Bacteriol. 184:4601-4611(2002).  
 CC -!- FUNCTION: Esterification, concomitant with transport, of exogenous long-chain fatty acids into metabolically active CoA thioesters for subsequent degradation or incorporation into phospholipids (By similarity).  
 CC -!- CATALYTIC ACTIVITY: ATP + a long-chain carboxylic acid + CoA = AMP + diphosphate + an acyl-CoA.  
 CC -!- COFACTOR: Magnesium (By similarity).  
 CC -!- SUBUNIT: Homodimer (Probable).  
 CC -!- SUBCELLULAR LOCATION: Partially membrane-associated (Potential).  
 CC -!- SIMILARITY: Belongs to the ATP-dependent AMP-binding enzyme family.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC EMBL; AJ414151; CAC90886.1; -  
 CC EMBL; AE013826; AAM85796.1; ALT\_INIT.  
 DR PIR; AB0253; AB0253.  
 DR InterPro; IPR000873; AMP-bind.  
 DR Pfam; PF00501; AMP-binding; 1.  
 DR PROSITE; PS00455; AMP BINDING; 1.  
 DR Ligase; Fatty acid metabolism; Magnesium; ATP-binding; Membrane; Complete proteome.  
 NP\_BIND 213 224 ATP (PROBABLE).  
 FT VARIANT 504 504 N -> K (IN STRAIN KIMS).  
 FT SEQUENCE 562 AA; 62639 MW; 812A872B3713D3FD CRC64;  
 SQ SEQUENCE 562 AA; 62639 MW; 812A872B3713D3FD CRC64;

Query Match 53.5%; Score 38; DB 1; Length 562;



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 26, 2004, 18:18:25 ; Search time 1.2769 Seconds  
(without alignments)  
3953.539 Million cell updates/sec

Title: US-09-302-812-14

Perfect score: 71

Sequence: 1 IALXLPNIXTQPIPL 16

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL 25:

- 1: sp\_archaea.\*
- 2: sp\_bacteria.\*
- 3: sp\_fungi.\*
- 4: sp\_human.\*
- 5: sp\_invertebrate.\*
- 6: sp\_mammal.\*
- 7: sp\_mhc.\*
- 8: sp\_organelle.\*
- 9: sp\_phage.\*
- 10: sp\_plant.\*
- 11: sp\_rodent.\*
- 12: sp\_virus.\*
- 13: sp\_vertebrate.\*
- 14: sp\_unclassified.\*
- 15: sp\_virus.\*
- 16: sp\_bacteriap.\*
- 17: sp\_archaeap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	65	91.5	920	11 Q8CB72	Q8CB72 mus musculus
2	65	91.5	961	11 Q80YQ6	Q80YQ6 mus musculus
3	65	91.5	968	11 Q88622	Q88622 mus musculus
4	65	91.5	972	11 Q9QYM2	Q9QYM2 rattus norv
5	65	91.5	976	4 Q9Y4W7	Q9Y4W7 homo sapien
6	65	91.5	976	4 Q86W56	Q86W56 homo sapien
7	65	91.5	976	4 Q7Z742	Q7Z742 homo sapien
8	63	91.5	977	6 Q02776	Q02776 bos taurus
9	46	64.8	723	5 Q960N8	Q960N8 drosophila
10	46	64.8	768	5 Q46043	Q46043 drosophila
11	44	62.0	317	5 Q16510	Q16510 caenorhabdi
12	43	60.6	524	5 Q16506	Q16506 caenorhabdi
13	43	60.6	1256	16 Q8YVT6	Q8YVT6 anabaena sp
14	42	59.2	203	16 Q8AAE5	Q8AAE5 bacteroides
15	42	59.2	249	5 Q8T306	Q8T306 tribolium c
16	42	59.2	513	12 Q83869	Q83869 narcissus 1

17	41.5	58.5	567	16	Q8A6R3	Q8A6R3 bacteroides
18	41	57.7	124	5	Q9XW97	Q9XW97 caenorhabdi
19	41	57.7	432	16	Q882S3	Q882S3 pseudomonas
20	41	57.7	499	16	Q82M67	Q82M67 streptomyce
21	41	57.7	1072	3	Q94537	Q94537 schizosacch
22	41	57.7	2096	16	Q7TWN7	Q7TWN7 mycobacteri
23	41	57.7	3157	16	Q50378	Q50378 mycobacteri
24	40	56.3	136	5	Q9NA23	Q9NA23 caenorhabdi
25	40	56.3	336	10	Q7XPK3	Q7XPK3 oryza sativ
26	40	56.3	368	10	Q8S164	Q8S164 oryza sativ
27	40	56.3	551	11	Q8VCA0	Q8VCA0 mus musculu
28	40	56.3	654	12	Q9DS87	Q9DS87 saint croix
29	40	56.3	774	5	Q9VJ74	Q9VJ74 drosophila
30	39	54.9	259	16	Q988L9	Q988L9 rhizobium 1
31	39	54.9	296	2	Q9ADW8	Q9ADW8 ehrlichia c
32	39	54.9	305	16	Q89V11	Q89V11 bradyrhizob
33	39	54.9	508	16	Q7UHI2	Q7UHI2 rhodospirall
34	39	54.9	526	10	Q9LZ26	Q9LZ26 arabidopsis
35	39	54.9	526	10	Q8GWI9	Q8GWI9 arabidopsis
36	39	54.9	585	16	Q8POA1	Q8POA1 xanthomonas
37	39	54.9	893	10	Q9SXH2	Q9SXH2 ipomoea bat
38	39	54.9	943	3	Q8X0B4	Q8X0B4 neurospora
39	39	54.9	997	16	Q8D4H3	Q8D4H3 vibrio vuln
40	39	54.9	1031	2	Q9AEQ8	Q9AEQ8 vibrio vuln
41	39	54.9	1031	2	Q9AHK2	Q9AHK2 vibrio vuln
42	39	54.9	1800	5	Q8IEI9	Q8IEI9 plasmodium
43	38	53.5	36	6	Q9GKJ9	Q9GKJ9 sus scrofa
44	38	53.5	138	1	Q9UXE2	Q9UXE2 sulfolobus
45	38	53.5	252	4	Q8N420	Q8N420 homo sapien

#### ALIGNMENTS

#### RESULT 1

Q8CB72 ID Q8CB72 PRELIMINARY; PRT; 920 AA.  
AC Q8CB72;  
DT 01-MAR-2003 (TREMREL. 23, Created)  
DT 01-MAR-2003 (TREMREL. 23, Last sequence update)  
DT 01-JUN-2003 (TREMREL. 24, Last annotation update)  
DE Poly.  
GN PARG.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
RX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6J; TISSUE=Bone;  
RX MEDLINE=22354683; PubMed=12466851;  
RA The FANTOM Consortium,  
RA the RIKEN Genome Exploration Research Group Phase I & II Team;  
RT \*Analysis of the mouse transcriptome based on functional annotation of  
RT 60,770 full-length cDNAs.\*;  
RL Nature 420:563-573 (2002).  
DR EMBL; AK036556; BAC29519.1;  
DR MGD; MGI:1347094; Parg.  
DR InterPro; IPR007724; PARG.  
DR Pfam; PF05028; PARG; 1.  
SQ SEQUENCE 920 AA; 103799 MW; 379E779E121B09A1 CRC64;

Query Match 91.5%; Score 65; DB 11; Length 920;

Best Local Similarity 87.5%; Pred. No. 0.0017; 2; Indels 0; Gaps 0;

Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 IALXLPNIXTQPIPL 16

Db 593 IALCLPNICTQPIPL 608

#### RESULT 2

Q80YQ6

Q9QYM2  
ID Q9QYM2  
AC Q9QYM2;  
PRELIMINARY;  
PRT; 972 AA.

Qy            1 IALXLPNIXTQPIPLL 16  
               |||||  
pb            609 IALCLPNICTOPIPLL 615

## 9. TILISA

Q86W56  
ID Q86W56 PRELIMINARY; PRT; 976 AA.  
AC Q86W56;  
DT 01-JUN-2003 (TREMBlrel. 24, Created)  
DT 01-JUN-2003 (TREMBlrel. 24, Last sequence update)  
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)  
DE Similar to poly (ADP-ribose) glycohydrolase.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Skin;  
RA Strausberg R.;  
RL Submitted (APR-2003) to the EMBL/GenBank/DBJ databases.  
DR EMBL; BC050560; AAH50560.1; -.  
DR GO; GO:0016787; F:hydrolase activity; IEA.  
DR InterPro; IPR007724; PARG.  
DR Pfam; PF05028; PARG; 1.  
KW Hydrolase.  
SQ SEQUENCE 976 AA; 111109 MW; D6646353C6D0180E CRC64;  
Query Match 91.5%; Score 65; DB 4; Length 976;  
Best Local Similarity 87.5%; Pred. NO. 0.0019;  
Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 1 IALXLPNIXTOPIPLL 16  
DB 600 IALCLPNICTOPIPLL 615  
RESULT 7  
Q7Z742  
ID Q7Z742 PRELIMINARY; PRT; 976 AA.  
AC Q7Z742;  
DT 01-OCT-2003 (TREMBlrel. 25, Created)  
DT 01-OCT-2003 (TREMBlrel. 25, Last sequence update)  
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)  
DE Hypothetical protein.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Testis;  
RA Strausberg R.L.; Feingold E.A.; Grouse L.H.; Derge J.G.;  
RA Klausner R.D.; Collins F.S.; Wagner L.; Shenmen C.M.; Schuler G.D.;  
RA Altschul S.F.; Zeeberg B.; Buetow K.H.; Schaefer C.F.; Bhat N.K.;  
RA Hopkins R.F.; Jordan H.; Moore T.; Max S.I.; Wang J.; Hsieh F.;  
RA Diatchenko L.; Marusina K.; Farmer A.A.; Rubin G.M.; Hong L.;  
RA Stapleton M.; Soares M.B.; Bonaldo M.F.; Casavant T.L.; Scheetz T.E.;  
RA Brownstein M.J.; Usdin T.B.; Toshiyuki S.; Carninci P.; Prange C.;  
RA Raha S.S.; Loquellano N.A.; Peters G.J.; Abramson R.D.; Mullahy S.J.;  
RA Bosak S.A.; McEwan P.J.; McKernan K.J.; Malek J.A.; Gunaratne P.H.;  
RA Richards S.; Worley K.C.; Hale S.; Garcia A.M.; Gay L.J.; Hulyk S.W.;  
RA Villalón D.K.; Muzny D.M.; Sodergren E.J.; Lu X.; Gibbs R.A.;  
RA Fahy J.; Helton E.; Kettman M.; Madan A.; Rodrigues S.; Sanchez A.;  
RA Whiting M.; Madan A.; Young A.C.; Shevchenko Y.; Bouffard G.G.;  
RA Blakesley R.W.; Touchman J.W.; Green E.D.; Dickson M.C.;  
RA Rodriguez A.C.; Grimwood J.; Schmutz J.; Myers R.M.; Butterfield Y.S.;  
RA Krzywinski M.I.; Skalska U.; Smailus D.E.; Schnerch A.; Schein J.E.;  
RA Jones S.J.; Marra M.A.;  
RA "Generation and initial analysis of more than 15,000 full-length human  
RT and mouse cDNA sequences."  
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Testis;  
RA Strausberg R.;  
RL Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.

DR EMBL; BC052966; AAH52966.1; -.  
KW Hypothetical protein.  
SQ SEQUENCE 976 AA; 111090 MW; 47C9EE1826C4A74A CRC64;  
Query Match 91.5%; Score 65; DB 4; Length 976;  
Best Local Similarity 87.5%; Pred. NO. 0.0019;  
Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 1 IALXLPNIXTOPIPLL 16  
DB 600 IALCLPNICTOPIPLL 615  
RESULT 8  
Q02776  
ID Q02776 PRELIMINARY; PRT; 977 AA.  
AC Q02776;  
DT 01-JUL-1997 (TREMBlrel. 04, Created)  
DT 01-JUL-1997 (TREMBlrel. 04, Last sequence update)  
DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)  
DE Poly (ADP-ribose) glycohydrolase.  
GN BPARG.  
OS Bos taurus (Bovine).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
OC Bovidae; Bovinae; Bos.  
OX NCBI\_TaxID=9913;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=97277328; PubMed=9115250;  
RA Lin W.; Ame J.C.; Aboul-Elia N.; Jacobson E.L.; Jacobson M.K.;  
RT "Isolation and characterization of the cDNA encoding bovine poly (ADP-  
RT ribose) glycohydrolase."  
RL J. Biol. Chem. 272:11895-11901 (1997).  
DR EMBL; U78975; AAH53370.1; -.  
DR GO; GO:0016787; F:hydrolase activity; IEA.  
DR InterPro; IPR007724; PARG.  
DR Pfam; PF05028; PARG; 1.  
KW Hydrolase.  
SQ SEQUENCE 977 AA; 110837 MW; 87D2100F979DF377 CRC64;  
Query Match 91.5%; Score 65; DB 6; Length 977;  
Best Local Similarity 87.5%; Pred. NO. 0.0019;  
Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 1 IALXLPNIXTOPIPLL 16  
DB 601 IALCLPNICTOPIPLL 616  
RESULT 9  
Q960N8  
ID Q960N8 PRELIMINARY; PRT; 723 AA.  
AC Q960N8;  
DT 01-DEC-2001 (TREMBlrel. 19, Created)  
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)  
DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)  
DE LD42380p.  
DR PARG OR EG:114E2.1 OR CG2864.  
GN Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7227;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Berkley;  
RA Stapleton M.; Brokstein P.; Hong L.; Agbayani A.; Carlson J.;  
RA Champe M.; Chavez C.; Dorsett V.; Farfan D.; Frise E.; George R.;  
RA Gonzalez M.; Guarin H.; Li P.; Liao G.; Miranda A.; Mungall C.J.;  
RA Nunoo J.; Pacleb J.; Paragas V.; Park S.; Phouanavong S.; Wan K.;  
RA Yu C.; Lewis S.E.; Rubin G.M.; Celniker S.;  
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.



Ame J.-C., Jacobson M.K.;  
"Isolation and characterization of the cDNA encoding Drosophila  
poly(ADP-ribose) glycohydrolase.";  
Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.  
[3]  
SEQUENCE OF 46-768 FROM N.A.  
Murphy L., Harris D., Barrell B.;  
Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.  
[4]  
SEQUENCE OF 46-768 FROM N.A.  
Benos P.;  
Submitted (DEC-1997) to the EMBL/GenBank/DBJ databases.  
EMBL; AF003428; AAF45886.1; -;  
EMBL; AF079556; AAC28734.1; -;  
EMBL; Z98254; CAB10913.1; -;  
FlyBase; FBgn0023216; Parg.  
GO; GO:0016787; F:hydrolase activity; IEA.  
InterPro; IPR007724; PARG.  
Pfam; PF05028; PARG; 1.  
KW Hydrolase. 768 AA; 86288 MW; 25620CA2B643FC04 CRC64;  
SQ SEQUENCE 768 AA; 86288 MW; 25620CA2B643FC04 CRC64;

Query Match 64.8%; Score 46; DB 5; Length 723;  
Best Local Similarity 50.0%; Pred. No. 5.3;  
Matches 8; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

OY 1 IALXLPNIXTQIPILL 16  
:|||||:  
DB 174 LALLRLPDLIQSFPVLL 199

RESULT 10  
ID O46043 PRELIMINARY; PRT; 768 AA.  
AC O46043;  
DT 01-JUN-1998 (TrEMBLrel. 06, Created)  
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
DE PARG protein.  
PARG OR EG:114E2.1 OR CG2864.  
GN Drosophila melanogaster (fruit fly).  
OS Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7227;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Berkely;  
RX MEDLINE=20196006; PubMed=10731132;  
RA Adams M.D., Celnik S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
Ananidides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,  
George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
Randon R.H., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,  
Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
Beeson K.Y., Bertram P.V., Bernier B.P., Brokstein P., Brotter P.,  
Borkova D., Botchan M.R., Bouck J., Brockton J., Cantarel D.A.,  
Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
Dodson K., Dou L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
Laško P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,  
Palazzo M., Pittman G.S., Pan S., Patterson J., Puri V., Reese M.G.,  
Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
Shue B.C., Sidon-Klamis I., Simpson M., Skupski M.P., Smith T.,  
Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
Swirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
Wang Z.-Y., Wasarman D.A., Weinstein G.M., Weissenbach J.,  
Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
Ye J., Yeh R.-F., Zaveri J.S., Zhao M., Zhang G., Zhou X., Zheng L.,  
Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu G., Zhu X., Smith H.O.,  
Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
"The genome sequence of Drosophila melanogaster";  
Science 287:2185-2195 (2000).  
[2]  
RP SEQUENCE FROM N.A.

AME J.-C., Jacobson M.K.;  
"Isolation and characterization of the cDNA encoding Drosophila  
poly(ADP-ribose) glycohydrolase.";  
Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.  
[3]  
SEQUENCE OF 46-768 FROM N.A.  
Murphy L., Harris D., Barrell B.;  
Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.  
[4]  
SEQUENCE OF 46-768 FROM N.A.  
Benos P.;  
Submitted (DEC-1997) to the EMBL/GenBank/DBJ databases.  
EMBL; AF003428; AAF45886.1; -;  
EMBL; AF079556; AAC28734.1; -;  
EMBL; Z98254; CAB10913.1; -;  
FlyBase; FBgn0023216; Parg.  
GO; GO:0016787; F:hydrolase activity; IEA.  
InterPro; IPR007724; PARG.  
Pfam; PF05028; PARG; 1.  
KW Hydrolase. 768 AA; 86288 MW; 25620CA2B643FC04 CRC64;  
SQ SEQUENCE 768 AA; 86288 MW; 25620CA2B643FC04 CRC64;

Query Match 64.8%; Score 46; DB 5; Length 768;  
Best Local Similarity 50.0%; Pred. No. 5.7;  
Matches 8; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

OY 1 IALXLPNIXTQIPILL 16  
:|||||:  
DB 219 LALLRLPDLIQSFPVLL 234

RESULT 11  
ID O16510 PRELIMINARY; PRT; 317 AA.  
AC O16510;  
DT 01-JAN-1998 (TrEMBLrel. 05, Created)  
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE CO3A7.13 protein.  
GN CO3A7.13.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditiida; Rhabditoidea;  
OC Rhabditiida; Peloderinae; Caenorhabditis.  
OX NCBI\_TaxID=6239;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Bristol N2;  
RX MEDLINE=94150718; PubMed=7906398;  
RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M., Coulson A.,  
Bonfield J., Burton J., Connell M., Cooley T., Cooper J., Coulson A.,  
Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,  
Gardner A., Green P., Kirsten J., Laister N., Latreille P.,  
Jones M., Kersey J., Kirwan J., Laister N., Jier M., Johnston L.,  
Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,  
Parsons J., Percy C., Rifkin L., Roopra A., Saunders D., Showkeen R.,  
Snaldon N., Smith A., Sonhammer E., Staden R., Sulston J.,  
Thierry-Wieg J., Thomas K., Vaubin M., Vaughan K., Waterston R.,  
Watson A., Weinstein L., Wilkinson-Sproat J., Wohlman P.,  
"2.2 Mb of contiguous nucleotide sequence from chromosome III of C.  
elegans.";  
Nature 368:32-38 (1994).  
[2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Bristol N2;  
RA Greco T., Bradshaw H., Elliott G.;  
"The sequence of C. elegans cosmid CO3A7.";  
Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.  
[3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Bristol N2;  
RA Waterston R.;  
Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.  
EMBL; AF016451; AAB65994.1; -;

Ame J.-C., Jacobson M.K.;  
"Isolation and characterization of the cDNA encoding Drosophila  
poly(ADP-ribose) glycohydrolase.";  
Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.  
[3]  
SEQUENCE OF 46-768 FROM N.A.  
Murphy L., Harris D., Barrell B.;  
Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.  
[4]  
SEQUENCE OF 46-768 FROM N.A.  
Benos P.;  
Submitted (DEC-1997) to the EMBL/GenBank/DBJ databases.  
EMBL; AF003428; AAF45886.1; -;  
EMBL; AF079556; AAC28734.1; -;  
EMBL; Z98254; CAB10913.1; -;  
FlyBase; FBgn0023216; Parg.  
GO; GO:0016787; F:hydrolase activity; IEA.  
InterPro; IPR007724; PARG.  
Pfam; PF05028; PARG; 1.  
Hydrolase. 768 AA; 86288 MW; 25620CA2B643FC04 CRC64;  
SEQUENCE

Query Match 64.8%; Score 46; DB 5; Length 723;  
Best Local Similarity 50.0%; Pred. No. 5.3;  
Matches 8; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

OY 1 IALXLPNIXTQIPILL 16  
:|||||:  
DB 174 LALLRLPDLIQSFPVLL 189

RESULT 10  
ID O46043 PRELIMINARY; PRT; 768 AA.  
AC O46043;  
DT 01-JUN-1998 (TrEMBLrel. 06, Created)  
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
DE PARG protein.  
PARG OR EG:114E2.1 OR CG2864.  
GN OS Drosophila melanogaster (fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Epherozoa; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7227;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Berkely;  
RX MEDLINE=20196006; PubMed=10731132;  
RA Adams M.D., Celnik S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
Anatitides P.G., Scher S.E., Li P.W., Hoskins R.A., Galie R.F.,  
George R.A., Lewis S.R., Richards S., Ashburner M., Henderson S.N.,  
Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
Randon R.H., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,  
Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
Abril J.F., Agayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,  
Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
Beeson K.Y., Bertram P.V., Bernier B.P., Brodtgen P., Brotter P.,  
Borkova D., Botchan M.R., Bouck J., Brockstein P., Brottier P.,  
Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
Hostin D., Houston K.A., Howland T.H., Ke Z., Kennison J.A., Ketchum K.A.,  
Khalil M., Kalush F., Karpen G.H., Kravitz S., Kulp D., Lai Z.,  
Kamel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
Laško P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,  
Palazzo M., Pittman G.S., Pan S., Patterson J., Puri V., Reese M.G.,  
Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
Shue B.C., Sidon-Klamis I., Simpson M., Skupski M.P., Smith T.,  
Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
Swirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissenbach J.,  
Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
Ye J., Yeh R.-F., Zaveri J.S., Zhao M., Zhang G., Zhao Q., Zheng L.,  
Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu G., Zhu X., Smith H.O.,  
Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
"The genome sequence of Drosophila melanogaster";  
Science 287:2185-2195 (2000).  
[2]  
SEQUENCE FROM N.A.

AME J.-C., Jacobson M.K.;  
"Isolation and characterization of the cDNA encoding Drosophila  
poly(ADP-ribose) glycohydrolase.";  
Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.  
[3]  
SEQUENCE OF 46-768 FROM N.A.  
Murphy L., Harris D., Barrell B.;  
Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.  
[4]  
SEQUENCE OF 46-768 FROM N.A.  
Benos P.;  
Submitted (DEC-1997) to the EMBL/GenBank/DBJ databases.  
EMBL; AF003428; AAF45886.1; -;  
EMBL; AF079556; AAC28734.1; -;  
EMBL; Z98254; CAB10913.1; -;  
FlyBase; FBgn0023216; Parg.  
GO; GO:0016787; F:hydrolase activity; IEA.  
InterPro; IPR007724; PARG.  
Pfam; PF05028; PARG; 1.  
Hydrolase. 768 AA; 86288 MW; 25620CA2B643FC04 CRC64;  
SEQUENCE

Query Match 64.8%; Score 46; DB 5; Length 723;  
Best Local Similarity 50.0%; Pred. No. 5.3;  
Matches 8; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

OY 1 IALXLPNIXTQIPILL 16  
:|||||:  
DB 219 LALLRLPDLIQSFPVLL 234

RESULT 11  
ID O16510 PRELIMINARY; PRT; 317 AA.  
AC O16510;  
DT 01-JAN-1998 (TrEMBLrel. 05, Created)  
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE CO3A7.13 protein.  
GN OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditoidea;  
OC Rhabditida; Peloderinae; Caenorhabditis.  
OX NCBI\_TaxID=6239;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Bristol N2;  
RX MEDLINE=94150718; PubMed=7906398;  
RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M., Coulson A.,  
Bonfield J., Burton J., Connell M., Cooley T., Cooper J., Coulson A.,  
Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,  
Gardner A., Green P., Kirsten J., Laister N., Latreille P.,  
Jones M., Kersey J., Kirwan J., Laister N., Jier M., Johnston L.,  
Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,  
Parsons J., Percy C., Rifkin L., Roopra A., Saunders D., Showkeen R.,  
Snaldon N., Smith A., Sonhammer E., Staden R., Sulston J.,  
Thierry-Wieg J., Thomas K., Vaubin M., Vaughan K., Waterston R.,  
Watson A., Weinstein L., Wilkinson-Sproat J., Wohlman P.,  
"2.2 Mb of contiguous nucleotide sequence from chromosome III of C.  
elegans.";  
Nature 368:32-38 (1994).  
[2]  
SEQUENCE FROM N.A.  
RC STRAIN=Bristol N2;  
RA Greco T., Bradshaw H., Elliott G.;  
"The sequence of C. elegans cosmid CO3A7.";  
Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.  
[3]  
SEQUENCE FROM N.A.  
RC STRAIN=Bristol N2;  
RA Waterston R.;  
Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.  
EMBL; AF016451; AAB65994.1; -;

```

DR PIR; T31897; T31897.
DR WormPep; C03A7.13; CB07871.
DR GO; GO:0005509; F:calcium ion binding; IEA.
DR InterPro; IPR002048; EF-hand.
DR PROSITE; PS00018; EF_HAND; 1.
SQ SEQUENCE 317 AA; 35452 MW; 8F9CB24E185DE488 CRC64;

Query Match 62.0%; Score 44; DB 5; Length 317;
Best Local Similarity 46.7%; Pred. No. 5.2;
Matches 7; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 1 IALXLPNIXTQPIPL 15
DB 173 VTGMLPNIASQDVPL 187

RESULT 12
ID O16506 PRELIMINARY; PRT; 524 AA.
AC O16506;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein C03A7.11.
GN C03A7.11.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RX MEDLINE=99069613; PubMed=9851916;
RA Waterston R.;
RT "Genome sequence of the nematode C. elegans: a platform for
investigating biology. The C. elegans Sequencing Consortium.";
RL Science 282:2012-2018 (1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RX Greco T., Bradshaw H., Elliott G.;
RA WormPep; C03A7.11; CE32564.
DR GO; GO:0016758; F:transferase activity, transferring hexosyl . . .; IEA.
DR GO; GO:0008152; P:metabolism; IEA.
DR InterPro; IPR002213; UDP_Gluco_trans.
DR Pfam; PF00201; UDPGT; 1.
KW Hypothetical protein.
SQ SEQUENCE 524 AA; 59372 MW; C69D9417434EF130 CRC64;

Query Match 60.6%; Score 43; DB 5; Length 524;
Best Local Similarity 46.7%; Pred. No. 14;
Matches 7; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 1 IALXLPNIXTQPIPL 15
DB 173 VTGMLPNIASQDVPL 187

RESULT 13
Q8YVT6 PRELIMINARY; PRT; 1256 AA.
ID Q8YVT6
AC Q8YVT6
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)

```

---

```

DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein A11888.
GN A11888.
OS Anabaena sp. (strain PCC 7120).
OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
OX NCBI_TaxID=103690;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21595285; PubMed=11759840;
RA Kaneko T., Nakamura Y., Wolk C.P., Kuritz T., Sasamoto S.,
RA Watanabe A., Iriquchi M., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
RA Nakazaki N., Shimpou S., Sugimoto M., Takazawa M., Yamada M.,
RA Yasuda M., Tabata S.;
RT "Complete genomic sequence of the filamentous nitrogen-fixing
cyanobacterium Anabaena sp. strain PCC 7120.";
RL DNA Res. 8:205-213 (2001).
DR EMBL; AF003587; BAB73587.1; -.
DR PIR; AB2042; AB2042.
DR InterPro; IPR008938; ARM.
DR InterPro; IPR007111; NACHT_NTPase.
DR Pfam; PF05729; NACHT; 1.
DR PROSITE; PS0837; NACHT; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 1256 AA; 147253 MW; DDFB5E19241BAC6 CRC64;

Query Match 60.6%; Score 43; DB 16; Length 1256;
Best Local Similarity 56.7%; Pred. No. 36;
Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 5 LPNIXTORIPL 16
DB 333 LDNVISQPIPL 344

RESULT 14
Q8AAE5 PRELIMINARY; PRT; 203 AA.
ID Q8AAE5
AC Q8AAE5;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Transcriptional regulator.
GN BT0520.
OS Bacteroides thetaiotaomicron.
OC Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
OC Bacteroidaceae; Bacteroides.
OX NCBI_TaxID=818;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=VPI-5482 / ATCC 29148;
RC MEDLINE=22550858; PubMed=12863928;
RA Xu J., Bjursell M.K., Hamrod J., Deng S., Carmichael L.K.,
RA Chiang H.C., Hooper L.V., Gordon J.I.;
RT "A genomic view of the human-Bacteroides thetaiotaomicron symbiosis.";
RL Science 299:2074-2076 (2003).
DR EMBL; AB016928; AAO75627.1; -.
DR GO; GO:0005622; C:intracellular; IEA.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0004672; F:protein kinase activity; IEA.
DR GO; GO:0003700; F:transcription factor activity; IEA.
DR GO; GO:000156; F:two-component response regulator activity; IEA.
DR GO; GO:0006468; P:protein amino acid phosphorylation; IEA.
DR GO; GO:0000160; P:two-component signal transduction system (p. . .; IEA.
DR InterPro; IPR00792; HTH_LuxR.
DR InterPro; IPR00719; Prot_kinase.
DR InterPro; IPR001789; Response_reg.
DR Pfam; PF0196; Gede; 1.
DR PRINTS; PR00038; HTHLUXR.
DR ProDom; PD000307; HTH_LuxR; 1.
DR ProDom; PD000001; Prot_kinase; 1.
DR ProDom; PD000039; Response_reg; 1.
DR SMART; SM00421; HTH_LuxR; 1.

```

DR PROSITE; PS00622; HTH\_LUXR\_FAMILY; 1.  
KW Complete proteome.  
SQ SEQUENCE 203 AA; 22763 MW; 4DC2F80D18D4F510 CRC64;  
Query Match 59.2%; Score 42; DB 16; Length 203;  
Best Local Similarity 66.7%; Pred. No. 7.7;  
Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 5 LPNIXTOPIPLL 16  
| | | | | | | |  
Db 29 LPNVKQPIELL 40  
| | | | | | | |

RESULT 15  
Q8T306 PRELIMINARY; PRT; 249 AA.  
AC Q8T306;  
DT 01-JUN-2002 (TrEMBLrel. 21, Created)  
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Helix-loop-helix transcription factor.  
GN HAIRY.  
OS Tribolium castaneum (Red flour beetle).  
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
OC Neoptera; Endopterygota; Coleoptera; Polyphaga; Cucujiformia;  
OC Tenebrionidae; Tribolium.  
OX NCBI\_TaxID=7070;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Eckert C., Wolff C., Wimmer E., Tautz D.;  
RT "Functional analysis of the regulatory region of the pair-rule gene hairy in Tribolium suggests regulatory divergence inspite of conserved expression.";  
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.  
CC -!- SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-HELIX (BHLH) FAMILY OF TRANSCRIPTION FACTORS.  
DR EMBL; AJ457831; CAD29886.1; -.  
DR GO; GO:0003677; F:DNA binding; IEA.  
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.  
DR InterPro; IPR001092; HLH\_basic.  
DR InterPro; IPR003650; Orange.  
DR Pfam; PF00010; HLH; 1.  
DR SMART; SM00353; HLH; 1.  
DR SMART; SM00511; ORANGE; 1.  
DR PROSITE; PS00038; HLH 1; 1.  
DR PROSITE; PS50888; HLH\_2; 1.  
SQ SEQUENCE 249 AA; 27252 MW; 364BB299EBEF429B CRC64;  
Query Match 59.2%; Score 42; DB 5; Length 249;  
Best Local Similarity 56.2%; Pred. No. 9.6;  
Matches 9; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

Qy 1 IALXLPNIXTOPIPLL 16  
| | | | | | | |  
Db 184 IALVLTQGSPLPLL 199  
| | | | | | | |

Search completed: May 26, 2004, 18:46:14  
Job time : 3.2769 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.  
OM protein - protein search, using sw model  
Run on: May 26, 2004, 17:20:49 ; Search time 4.80653 Seconds  
(without alignments)  
3249.964 Million cell updates/sec  
Title: US-09-302-812-4\_COPY\_677\_977  
Perfect score: 1597  
Sequence: 1 KTLFCYFRVTEKKTGLVT.....YHAVSCAETADHSGQRTGT 300  
Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5  
Searched: 141681 seqs, 52070155 residues  
Total number of hits satisfying chosen parameters: 141681  
Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries  
Database : SwissProt\_42.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	91.5	5.7	261	1	LAS1 HUMAN
2	91.5	5.7	491	1	K2M2 SHEEP
3	90.5	5.7	263	1	LAS1 MOUSE
4	89	5.6	727	1	IF2M MOUSE
5	86	5.4	153	1	ENPP BPP21
6	86	5.4	153	1	ENPP LAMBD
7	85	5.3	523	1	RPB2 HALN1
8	84.5	5.3	304	1	CYSD_XYLFA
9	84	5.3	2298	1	KCF2_L0TFA
10	83.5	5.2	153	1	RZPD_ECOLI
11	83	5.2	1135	1	RBL2 RAT
12	82.5	5.2	624	1	HTPG_CLOAB
13	82.5	5.2	854	1	AKA3 BOVIN
14	82.5	5.2	998	1	BGAL_LACIA
15	81.5	5.1	982	1	CHLB_HUMAN
16	81	5.1	1575	1	IGG2 HUMAN
17	79.5	5.0	618	1	IORA METH
18	79	4.9	721	1	HS90 THEPA
19	78.5	4.9	317	1	YCQ5 YEAST
20	78.5	4.9	514	1	TSPP_HUMAN
21	78.5	4.9	761	1	EPG1 YEAST
22	78.5	4.9	989	1	IDB DROME
23	78	4.9	405	1	SRCY ANTSP
24	78	4.9	469	1	ORB6 SCHPO
25	78	4.9	703	1	HS83 TRYBB
26	77.5	4.9	252	1	HIS6 RHOC
27	77.5	4.9	283	1	YR34 DEIFA
28	77.5	4.9	314	1	MAG3 HUMAN
29	77.5	4.9	384	1	ISCF BRUME
30	77.5	4.9	404	1	YXID TREPA
31	77.5	4.9	569	1	YXID BACSU
32	77	4.8	344	1	FLIM TREPA
33	77	4.8	414	1	YF24_MYCTU

34	76.5	4.8	314	1	MAG6 HUMAN
35	76.5	4.8	343	1	LPSD RHIME
36	76.5	4.8	477	1	NIFD_WETMP
37	76.5	4.8	520	1	UBF3_MOUSE
38	76.5	4.8	649	1	RA32_SCHPO
39	76	4.8	366	1	HI82_BACCR
40	76	4.8	593	1	CDR1_SCHPO
41	76	4.8	627	1	PLSL_HUMAN
42	76	4.8	657	1	UV88_CAMJE
43	76	4.8	837	1	HFC3_HAEIN
44	76	4.8	839	1	OS94_HUMAN
45	76	4.8	4447	1	PK3K_BACSU

P43360	homo sapien
Q9r9n0	rhizobium m
P71526	methanococc
Q91w36	mus musculus
Q09683	schizosacch
Q81c43	bacillus ce
P07334	schizosacch
P13796	homo sapien
Q9pm7	campylobact
P45998	haemophilus
Q95757	homo sapien
P40803	bacillus su

ALIGNMENTS

RESULT 1  
LAS1 HUMAN  
ID LAS1 HUMAN STANDARD; PRT; 261 AA.  
AC Q14847; Q96ED2;  
DT 15-JUL-1998 (Rel. 36, Created)  
DT 15-JUL-1998 (Rel. 36, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE LIM and SH3 domain protein 1 (LASP-1) (MLN 50).  
GN LASP1 OR MLN50.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Breast carcinoma;  
RX MEDLINE=96039245; PubMed=7490069;  
RA Tomasetto C., Regnier C., Moog-Lutz C., Mattei M.-G., Chenard M.-P.,  
RA Lidereau R., Basset P., Rio M.-C.;  
RT "Identification of four novel human genes amplified and overexpressed  
RT in breast carcinoma and localized to the q11-q21.3 region of  
RT chromosome 17.";  
RL Genomics 28:367-376(1995).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Liver;  
RX MEDLINE=22388257; PubMed=12477932;  
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heish F.,  
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
RA Brownstein M.J., Uadin T.B., Toohiyuki S., Carninci P., Prange C.,  
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,  
RA Boak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Fahy J., Helton E., Kettner M., Madan A., Rodriguez S., Sanchez A.,  
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
RA Butterfield Y.C., Krzywinski M.I., Skalska U., Smalusz D.E.,  
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
RT "Generation and initial analysis of more than 15,000 full-length  
RT human and mouse cDNA sequences.";  
RT Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
RN [3]  
RP DOMAINS.  
RX MEDLINE=96033982; PubMed=7589475;  
RA Tomasetto C., Moog-Lutz C., Regnier C.H., Schreiber V., Basset P.,  
RA Rio M.-C.;  
RT "Lasp-1 (MLN 50) defines a new LIM protein subfamily characterized by  
RT the association of LIM and SH3 domains.";  
RL FEBS Lett. 373:245-249(1995).  
CC -!- SIMILARITY: Contains 1 LIM zinc-binding domain.

```

CC -!- SIMILARITY: Contains 1 SH3 domain.
CC -!- DATABASE: NAME-Atlas Genet. Cytogenet. Oncol. Haematol.;
CC WWW="http://www.infobiogen.fr/services/chronocancer/Genes/Laspid203.html".
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X82456; CAA57833.1; -.
CC EMBL; BC012460; AAH12460.1; -.
CC HSSP; P80171; 12FO.
CC SWISS-2DPAGE; Q14847; HUMAN.
CC Genew; HGNC:6513; LASP1.
CC MIM; 602920; -.
CC GO; GO:0005070; F:SH3/SH2 adaptor protein activity; TAS.
CC InterPro; IPR001781; LIM.
CC InterPro; IPR000900; Nebulin.
CC InterPro; IPR001452; SH3.
CC Pfam; PF00412; LIM; 1.
CC Pfam; PF00880; Nebulin; 2.
CC Pfam; PF00018; SH3; 1.
CC PRINTS; PR00452; SH3DOMAIN.
CC ProDom; PD000094; LIM; 1.
CC ProDom; PD000066; SH3; 1.
CC SMART; SM00132; LIM; 1.
CC SMART; SM00227; NEBU; 2.
CC SMART; SM00326; SH3; 1.
CC PROSITE; PS00478; LIM_DOMAIN_1; 1.
CC PROSITE; PS00023; LIM_DOMAIN_2; 1.
CC PROSITE; PS00002; SH3; 1.
CC LIM domain; Metal-binding; Zinc; SH3 domain.
CC FT DOMAIN 5 56 LIM.
CC FT DOMAIN 202 261 SH3.
CC FT DOMAIN 201 204 POLY-GLY.
CC FT CONFLICT 79 79 E -> R (IN REF. 2).
CC FT CONFLICT 210 210 V -> A (IN REF. 2).
CC FT CONFLICT 220 220 E -> A (IN REF. 2).
CC SQ SEQUENCE 261 AA; 29717 MW; 3989B98605B3639 CRC64;

Query Match 5.7%; Score 91.5; DB 1; Length 261;
Best Local Similarity 22.6%; Pred. No. 0.72;
Matches 53; Conservative 32; Mismatches 80; Indels 69; Gaps 9;

Qy 12 EKKPTGLVTFTRQS---LEDFPEWERC-EKPLTRLHVTYEGTIEEN-GQGMLQVDFANRP 66
Db 48 EKKYCNAHYPKQSFVTWADTPENLRKQOSELOSQVRYKEEFKNKGK----- 97

Qy 67 VGGVTSAGLVQEBIRFLINPELIISRLFTVLVDHNECLITGTGEQYSEYTGAYETYRWS 126
Db 98 -----FSVVADTPELRIKKTQDQISNLIKHYEEFEKS 129

Qy 127 RSHEDGSDRDCRCRCEIVEIDALHFRYLDQVFP-----EKMRRLNKAYCG 175
Db 130 RMGPSGEGMEPERDSQ---DGSYRRPLEQQPHIPTSAFYQPOQPOQVQSYGG 185

Qy 176 FLRPGVSENLSAVATGNWCGAGFGDARLKALIQILAAARDDVVVFTFGDS 229
Db 186 YKEP-AAPVSIQRSAPG-----GGKRYRAVDY-----SAADEVSVFQGD 228

RESULT 2
K2M2 SHEEP STANDARD; PRT; 491 AA.
AC P15241;
DT 01-APR-1990 (Rel. 14, Created)
DT 16-APR-1990 (Rel. 14, Last sequence update)
DT 01-OCT-2001 (Rel. 40, Last annotation update)
DE Keratin, type II microfibrillar, component 7C.
OS Ovis aries (Sheep).

```

```

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Caprinae; Ovis.
OC NCBI_TaxID=9940;
RN [1]
RP SEQUENCE
RX MEDLINE=90026244; PubMed=2803231;
RA Sparrow L.G., Robinson C.P., McMahon D.T.W., Rubira M.R.;
RT "The amino acid sequence of component 7c, a type II intermediate-
RL filament protein from wool.";
RL Biochem. J. 261:1015-1022(1989).
CC -!- FUNCTION: WOOL MICROFIBRILLAR KERATIN.
CC -!- MISCELLANEOUS: THERE ARE TWO TYPES OF CYTOSKELETAL AND
CC MICROFIBRILLAR KERATIN, I AND II (40-55 AND 56-70 KILODALTONS,
CC RESPECTIVELY). AT LEAST ONE TYPE OF CYTOSKELETAL KERATIN IS
CC PRESENT IN ALL VERTEBRATE EPITHELIAL CELLS.
CC -!- MISCELLANEOUS: THE LOW-SULFUR PROTEINS, DERIVED FROM THE
CC MICROFIBRILLAR FRACTION OF WOOL EXTRACTS, ARE COMPOSED OF TWO
CC FAMILIES, EACH CONSISTING OF 4 HOMOLOGOUS SUBUNITS: THE TYPE I
CC COMPONENTS (8C-1, 8C-2, 8A & 8B) AND THE TYPE II COMPONENTS (5,
CC 7A, 7B, AND 7C).
CC -!- SIMILARITY: Belongs to the intermediate filament family.
CC InterPro; IPR001664; IF.
CC InterPro; IPR003054; Keratin_II.
CC Pfam; PF00038; filament; 1.
CC PRINTS; PR01276; TYPE2KERATIN.
CC PROSITE; PS00226; IF; 1.
CC Intermediate filament; Coiled coil; Keratin.
CC MOD RES 1 1 BLOCKED.
CC FT DOMAIN 1 109 HEAD.
CC FT DOMAIN 110 416 ROD.
CC FT DOMAIN 417 491 TAIL.
CC FT DOMAIN 110 144 COIL 1A.
CC FT DOMAIN 145 154 LINKER 1.
CC FT DOMAIN 155 255 COIL 1B.
CC FT DOMAIN 256 272 LINKER 12.
CC FT DOMAIN 273 416 COIL 2.
CC FT VARIANT 74 74 C -> G OR S.
CC FT VARIANT 80 80 C -> S.
CC FT VARIANT 144 144 F -> Y.
CC FT VARIANT 232 232 S -> V.
CC FT VARIANT 276 276 C -> D OR N.
CC FT VARIANT 284 284 Q -> H.
CC FT UNSURE 1 2 CG -> GC.
CC SQ SEQUENCE 491 AA; 53681 MW; A801771FE3831ABE CRC64;

Query Match 5.7%; Score 91.5; DB 1; Length 491;
Best Local Similarity 23.9%; Pred. No. 1.6;
Matches 62; Conservative 32; Mismatches 96; Indels 69; Gaps 13;

Qy 50 IEENGQGMLO-VDFANRFVGGVTSAGLVQEBIRFL---INPELLISRLFTVLVDHNECL 105
Db 228 LEANSEALIQEIDFLRR-----LYQEBIRVLQANISDTSVIVKMDNSRDLNMDCI 277

Qy 106 IITGTEQYSEYT--GYAETYRWSRSHEDGSDRDCRCEIVEIDALH---FRYLDQF 160
Db 278 VAEIKAQYDDIASRSRAEBSWYRS-----KCEIKATVTRHGETLRRTKEEI 325

Qy 161 -----VPEKMRRLNKAYC--GFLRPGVS--SENL SAVATGNWCGAGFGDARLKALIQIL 212
Db 326 NELNRVQLRLTAEVENAKCNKSLAAVTAQEQGEVALNDARCKLAGLEALQKAKQDM 385

Qy 213 AAAAARDVVYFTFGDSELMRDYSHMIFLTERKLTVDGVYKLLLRYYNECRNCSTPGP 272
Db 386 ACLIKE-----YQEVMSKGLDIE-----IATYRRLLEG-----BEQRLCEGVG- 425

Qy 273 DIKLYPIYHVESCAETA 291
Db 426 -----AVNVCVSSS 434

RESULT 3
LASI_MOUSE

```

ID LAS1 MOUSE STANDARD; PRT; 263 AA.  
 AC Q61732; Q62416;  
 DT 15-JUL-1998 (Rel. 36, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DT 15-MAR-2004 (Rel. 43, Last annotation update)  
 GN LIM and SH3 domain protein 1 (LASP-1) (MLN 50).  
 GE LASP1 OR MLN50.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 ON NCBI\_TaxID=10090;  
 RN [1]  
 RP MEDLINE=98172750; PubMed=9511759;  
 RA Bio M.-C.;  
 RA "Chromosomal assignment and expression pattern of the murine Lasp-1  
 RT gene.";  
 RL Gene 207:171-175(1998).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Mammary gland;  
 RX MEDLINE=22388257; PubMed=12477932;  
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
 RA Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C.,  
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,  
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahey J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,  
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
 RA Butterfield Y.S.N., Krzywinski M.I., Skalek U., Smailus D.E.,  
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length  
 RL human and mouse cDNA sequences.";  
 RN Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 RN [3]  
 RP SEQUENCE OF 59-263 FROM N.A.  
 RX MEDLINE=98294438; PubMed=9630982;  
 RA Sparks A.B., Hoffman N.G., McConnell S.J., Fowlkes D.M., Kay B.K.;  
 RA "Cloning of ligand targets: systematic isolation of SH3 domain-  
 RT containing proteins.";  
 RL Nat. Biotechnol. 14:741-744(1996).  
 CC -!- SIMILARITY: Contains 1 LIM zinc-binding domain.  
 CC -!- SIMILARITY: Contains 1 SH3 domain.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; X96973; CAA65659.1; -;  
 CC EMBL; BC010840; AAH10840.1; -;  
 CC EMBL; U58882; AAC52639.1; -;  
 CC HSSP; P80171; 1ZFO  
 CC MGD; MGI:105656; Laspl.  
 CC InterPro; IPR001781; LIM.  
 CC InterPro; IPR000900; Nebulin.  
 CC InterPro; IPR001452; SH3.  
 CC Pfam; PF00412; LIM; 1.  
 CC Pfam; PF00880; Nebulin; 2.  
 CC Pfam; PF00018; SH3; 1.  
 CC PRINTS; PR00452; SH3DOMAIN.

DR ProDom; PD000094; LIM; 1.  
 DR ProDom; PD000066; SH3; 1.  
 DR SMART; SM00132; LIM; 1.  
 DR SMART; SM00227; NEBU; 2.  
 DR SMART; SM00326; SH3; 1.  
 DR PROSITE; PS00478; LIM DOMAIN 1; 1.  
 DR PROSITE; PS0023; LIM DOMAIN 2; 1.  
 DR PROSITE; PS0002; SH3; 1.  
 KW LIM domain; Metal-binding; Zinc; SH3 domain.  
 FT DOMAIN 5 56 LIM.  
 FT DOMAIN 204 263 SH3.  
 FT DOMAIN 174 180 POLY-GLN.  
 FT DOMAIN 203 206 POLY-GLY.  
 SQ SEQUENCE 263 AA; 29994 MW; A6CA2FC2E451433E CRC64;  
 Query Match 5.7%; Score 90.5; DB 1; Length 263;  
 Best Local Similarity 22.5%; Pred. No. 0.89;  
 Matches 53; Conservative 31; Mismatches 81; Indels 71; Gaps 9;  
 QY 12 EKKTGLVTFTROS---LEDFPEWERC-EKPLTRHLVHYEGTTEEN-GQMLQVDFANRF 66  
 DB 48 EKFPYCNHYPKQSFPTWADTPENLKQSELSQVRYKEEFKKKGK----- 97  
 QY 67 VGGVTSAGLVQDEIRFLINPELIIISRLFTVLDHNECLITGTQYSEYTYGYAIVRWS 126  
 DB 98 -----FSWVADTPELQRIKKTQDQISNIKYHEEFKS 129  
 QY 127 RSHEDGSRDCCERCCTEIVADALHFRYLDQFVP-----EKMRELNKAY 173  
 DB 130 RMGPSGEGVEPERR---EAQDSSYRRPTEQQQPHHPTSPAPYQPOQOWTSSY 185  
 QY 174 CGFLRPGVSENLAVATGNWCGAGFGDARLKALQILAAARDDVVVYFTFGDS 229  
 DB 186 GGYKEP-APVSIQISAPG-----GGCKRYAVVDY---SHADEVSFQDGT 230  
 RESULT 4  
 IF2M MOUSE STANDARD; PRT; 727 AA.  
 ID IF2M MOUSE  
 AC Q913J5;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Translation initiation factor IF-2, mitochondrial precursor (IF-2mt)  
 DE (IF-2(Mt)).  
 GN MTIF2.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 ON NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22388257; PubMed=12477932;  
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
 RA Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C.,  
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,  
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahey J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,  
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
 RA Butterfield Y.S.N., Krzywinski M.I., Skalek U., Smailus D.E.,  
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length  
 RL human and mouse cDNA sequences.";  
 RN Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).

-- FUNCTION: One of the essential components for the initiation of protein synthesis. Protects formylmethionyl-tRNA from spontaneous hydrolysis and promotes its binding to the 30S ribosomal subunits. Also involved in the hydrolysis of GTP during the formation of the 70S ribosomal complex (By similarity).

-- SUBUNIT: Monomer (By similarity).

-- SUBCELLULAR LOCATION: Mitochondrial (By similarity).

-- SIMILARITY: Belongs to the IP-2 family.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL; BC016590; AAH16590.1; --  
MGI; MGI:1924034; Mif2.  
InterPro; IPR000795; EF\_GTPbind.  
InterPro; IPR004161; IF2.  
InterPro; IPR000178; IF2.  
InterPro; IPR005225; Small\_GTP.  
InterPro; IPR009000; Translat\_factor.  
Pfam; PF00009; GTP\_EFTU; 1.  
Pfam; PF03144; GTP\_EFTU\_D2; 1.  
ProDom; PD186100; IF2; 1.  
TIGRFAMs; TIGR00487; IP-2; 1.  
TIGRFAMs; TIGR00231; small\_GTP; 1.  
PROSITE; PS01176; IF2; 1.  
Initiation factor; Protein biosynthesis; GTP-binding; Transit peptide;  
KW Mitochondrion.  
FT TRANSIT 1 29 MITOCHONDRION (POTENTIAL).  
FT CHAIN 30 727 TRANSLATION INITIATION FACTOR IP-2.  
FT DOMAIN 181 330 G-DOMAIN.  
FT NP\_BIND 187 194 GTP (BY SIMILARITY).  
FT NP\_BIND 234 237 GTP (BY SIMILARITY).  
FT NP\_BIND 288 291 GTP (BY SIMILARITY).  
SQ SEQUENCE 727 AA; 81346 MW; A8C3FF8646A45F6D CRC64;

Query Match 5.6%; Score 89; DB 1; Length 727;  
Best Local Similarity 20.7%; Pred. No. 4.4;  
Matches 41; Conservative 34; Mismatches 63; Indels 60; Gaps 10;

QY 121 EYRWSRS-----HEGSDRDCERCTHIVALDHP--RYLDQFVPEK 164  
DB 442 EVIEWRKSEQKEKGKDLKIMEKREHQAHRKARE--KYGSLHWKERSYI-KFLERK 498  
QY 165 MRRELKAYCGPLRPGVSEN---LSAVATGCGGAFGGDARLKALIQIL---AAAA 216  
DB 499 QORP-----LKPKEKVRQSNVLPITIKGD-----VDGSEALINLDTVDASHE 543  
QY 217 AERDVVYTFGDSSELMRDYISMHIFLTKRTVGDVYKLLRYNEEC-----RNCSTPG 271  
DB 544 CELELVHFGLED-----ISENDVTFATFDGVIYGFNVGSAIQQAQKG 590  
QY 272 PDIKLYPIFYHVESCAE 289  
DB 591 VKIKLHKIYHLEDLQEQ 608

RESULT 5  
ENPP\_BPP21  
ID ENPP\_BPP21 STANDARD; PRT; 153 AA.  
AC P27358;  
DT 01-AUG-1992 (Rel. 23, Created)  
DT 01-AUG-1992 (Rel. 23, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Endopeptidase (EC 3.4.-.-) (Lysis protein R2).  
GN RZ.  
OS Bacteriophage P21 (Bacteriophage 21).  
OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Siphoviridae;  
OC Lambda-like viruses.

Query Match 5.4%; Score 86; DB 1; Length 153;  
Best Local Similarity 23.3%; Pred. No. 1.1;  
Matches 42; Conservative 24; Mismatches 56; Indels 58; Gaps 8;

QY 104 CLITTEQYSEYTGTAETRWRSRSHEDGSRD-----DCERRCTEIVADALHFR 154  
DB 14 CIIVCLSWAVNHYRDNALTAKAQR---DKNARELTLANRVITDIQMRQDVAALDAKYTK 70  
QY 155 RYLDQFVPEKM-----RRELN-KAYCGPLRPGVSENLSAVATGCGGAFGGDARL 205  
DB 71 ELADAKAENDALRDDVAAGRRRLHKAVCQSVREATTASGVDNAAS-----PRL 119  
QY 206 KALIQILAAAAERDVVYTFGDSSELMRDYISMHIFLTKRTVGDVYKLLRYNEECR 265  
DB 120 -----ADTAERD--YFT-----LREELITMQKLEGTKYNEQCR 153

RESULT 6  
ENPP\_LAMB  
ID ENPP\_LAMB STANDARD; PRT; 153 AA.  
AC P00726;  
DT 21-JUL-1986 (Rel. 01, Created)  
DT 21-JUL-1986 (Rel. 01, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Endopeptidase (EC 3.4.-.-) (Lysis protein R2).  
GN RZ.  
OS Bacteriophage lambda.  
OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Siphoviridae;  
OC Lambda-like viruses.  
OC NCBI\_TaxID=10710;  
RW [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=83189071; PubMed=6221115;  
RA Sanger F., Coulson A.R., Hong G.F., Hill D.F., Petersen G.B.;  
RL "Nucleotide sequence of bacteriophage lambda DNA.";  
RL J. Mol. Biol. 162:729-773(1982).  
CC -- FUNCTION: Necessary for host cell lysis. It is believed to code for an endopeptidase that cleaves the amino-carboxyl cross-link between the diaminopimelic acid and D-alanine residues in the murein component of the bacterial cell wall.  
CC -- SIMILARITY: Belongs to peptidase family U8.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way

NCBI\_TaxID=10711;  
[1]  
SEQUENCE FROM N.A.  
RX MEDLINE=91210180; PubMed=2019562;  
RA Bonovich M.T., Young R.;  
RL "Dual start motif in two lambdoid S genes unrelated to lambda S.";  
RL J. Bacteriol. 173:2897-2905(1991).  
CC -- FUNCTION: Necessary for host cell lysis. It is believed to code for an endopeptidase that cleaves the amino-carboxyl cross-link between the diaminopimelic acid and D-alanine residues in the murein component of the bacterial cell wall.  
CC -- SIMILARITY: Belongs to peptidase family U8.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL; M65239; AAA32351.1; --  
InterPro; IPR004929; Phage\_lysis.  
Pfam; PF03245; Phage\_lysis; 1.  
KW Hydrolase; Protease; Bacteriolytic enzyme.  
SQ SEQUENCE 153 AA; 17325 MW; 3291A79D77AA1EFD CRC64;

Query Match 5.4%; Score 86; DB 1; Length 153;  
Best Local Similarity 23.3%; Pred. No. 1.1;  
Matches 42; Conservative 24; Mismatches 56; Indels 58; Gaps 8;

QY 104 CLITTEQYSEYTGTAETRWRSRSHEDGSRD-----DCERRCTEIVADALHFR 154  
DB 14 CIIVCLSWAVNHYRDNALTAKAQR---DKNARELTLANRVITDIQMRQDVAALDAKYTK 70  
QY 155 RYLDQFVPEKM-----RRELN-KAYCGPLRPGVSENLSAVATGCGGAFGGDARL 205  
DB 71 ELADAKAENDALRDDVAAGRRRLHKAVCQSVREATTASGVDNAAS-----PRL 119  
QY 206 KALIQILAAAAERDVVYTFGDSSELMRDYISMHIFLTKRTVGDVYKLLRYNEECR 265  
DB 120 -----ADTAERD--YFT-----LREELITMQKLEGTKYNEQCR 153

RESULT 6  
ENPP\_LAMB  
ID ENPP\_LAMB STANDARD; PRT; 153 AA.  
AC P00726;  
DT 21-JUL-1986 (Rel. 01, Created)  
DT 21-JUL-1986 (Rel. 01, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Endopeptidase (EC 3.4.-.-) (Lysis protein R2).  
GN RZ.  
OS Bacteriophage lambda.  
OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Siphoviridae;  
OC Lambda-like viruses.  
OC NCBI\_TaxID=10710;  
RW [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=83189071; PubMed=6221115;  
RA Sanger F., Coulson A.R., Hong G.F., Hill D.F., Petersen G.B.;  
RL "Nucleotide sequence of bacteriophage lambda DNA.";  
RL J. Mol. Biol. 162:729-773(1982).  
CC -- FUNCTION: Necessary for host cell lysis. It is believed to code for an endopeptidase that cleaves the amino-carboxyl cross-link between the diaminopimelic acid and D-alanine residues in the murein component of the bacterial cell wall.  
CC -- SIMILARITY: Belongs to peptidase family U8.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way

```

RESULT 7
RPB2=HALN1
ID RPB2 HALN1 STANDARD; PRT; 523 AA.
AC F15352; Q9HM77;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE DNA-directed RNA polymerase subunit B' (EC 2.7.7.6).
GN RPB2 OR VNC2666G.
OS Halobacterium sp. (strain NRC-1 / ATCC 700922 / JCM 11081), and
OS Halobacterium halobium.
OC Archaea; Euryarchaeota; Halobacteria; Halobacteriales;
OC Halobacteriaceae; Halobacterium.
OX NCBI_TaxID=64091, 2242;
[1]
RN
SEQUENCE FROM N.A.
RC STRAIN=NRC-1;
RC MEDLINE=20504483; PubMed=11016950;
RX Ng W.V., Kennedy S.P., Mahairas G.G., Berquist B., Pan M.,
RX Shukla H.D., Laaky S.R., Baliga N.S., Thorsson V., Shroga J.,
RX Swartzell S., Weir D., Hall J., Dahl T.A., Weiti R., Goo Y.A.,
RX Leitzauser B., Keller K., Cruz R., Danson M.J., Hough D.W.,
RX Maddocks D.G., Jablonksi P.E., Krebs M.P., Angevine C.M., Dale H.,
RX Isenbarger T.A., Peck R.F., Pohlschroder M., Spudich J.L., Jung K.-H.,
RX Alam M., Freitas T., Hou S., Daniels C.J., Dennis P.P., Omer A.D.,
RX Ebhardt H., Lowe T.M., Liang P., Riley M., Hood L., DasSarma S.;
RT "Genome sequence of Halobacterium species NRC-1.";
RL Proc. Natl. Acad. Sci. U.S.A. 97:12176-12181(2000).
[2]
RN
SEQUENCE FROM N.A., AND SEQUENCE OF 1-17.
RC SPECIES=H.halobium; STRAIN=R1;
RX MEDLINE=89199633; PubMed=2495365;
RX Leffers H., Gropp F., Lottspeich F., Zillig W., Garrett R.A.;
RT "Sequence, organization, transcription and evolution of RNA
RT polymerase subunit genes from the archaeobacterial extreme halophiles
RT Halobacterium halobium and Halococcus morrhuae.";
RL J. Mol. Biol. 206:1-17(1989).
CC -1- FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION
CC OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS
CC SUBSTRATES. THE B (B'+B' AND BETA) SUBUNITS HAVE BEEN IMPLICATED
CC IN DNA PROMOTER RECOGNITION AND ALSO IN NUCLEOTIDE BINDING.
CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
CC {RNA} (N).
CC -1- COFACTOR: Zinc.
CC -1- SUBUNIT: This RNA polymerase is composed of 5 large subunits:

```



RA Facinani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,  
 RA Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furlan L.R.,  
 RA Garnier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,  
 RA Ho P.L., Hoheisel J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,  
 RA Krieger J.E., Kuranae E.E., Laigret F., Lambais M.R., Leite L.C.C.,  
 RA Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,  
 RA Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Marino C.L.,  
 RA Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,  
 RA Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,  
 RA Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,  
 RA Nhani A. Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,  
 RA de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,  
 RA Peixoto B.R., Pereira G.A.G., Pereira H.A. Jr., Pesquero J.B.,  
 RA Paquaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,  
 RA de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,  
 RA da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,  
 RA da Silveira J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,  
 RA de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tsubako M.H.,  
 RA Vallada H., van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,  
 RA Zago M.A., Zatz M., Meidanis J., Setubal J.C.;  
 RL "The genome sequence of the plant pathogen *Xylella fastidiosa*;"  
 Nature 406:151-159(2000).  
 CC -!- CATALYTIC ACTIVITY: ATP + sulfate = diphosphate + adenylylsulfate.  
 CC -!- PATHWAY: First step in the sulfate activation pathway. This  
 CC reaction occurs early in the reductive branch of the cysteine  
 CC biosynthetic pathway.  
 CC -!- SUBUNIT: Heterodimer composed of cysD, the smaller subunit, and  
 CC cysN (By similarity).  
 CC -!- SIMILARITY: Belongs to the PAPS reductase family. CysD subfamily.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; AE003980; FAF94309.1; ALT\_INIT.  
 CC PIR; F82672.  
 CC HAMAP; MF 00064; -. 1.  
 CC InterPro; IPR002500; PAPS reduct.  
 CC Pfam; PF01507; PAPS\_reduct; 1.  
 CC Transferrase; Nucleotidyltransferase; Cysteine biosynthesis;  
 CC Complete proteome.  
 CC SEQUENCE 304 AA; 34886 MW; F4B1EA01BD81EE2 CRC64;  
 Query Match 5.3%; Score 84.5; DB 1; Length 304;  
 Best Local Similarity 22.6%; Pred. No. 3.7;  
 Matches 62; Conservative 32; Mismatches 97; Indels 83; Gaps 14;  
 QY 38 PLTRLHYTYEGTIBENGQMLQVDFANRFVGGVTSAGLVOEIRFLINPELLI---ISRL 94  
 Db 59 PIPLLHVDTRMKFRE-----MITFRD---RYAETGV---QLRVHINPGVAQINPI 105  
 QY 95 FTEVLDDNECLIIITGEQSYTYGTYAETYRWSRSHEDGSDCERCCTEIVADLHPR 154  
 Db 106 THGAHVHTDVNKTQGLRQALQGGQFDAAI-----GGARDEKRSRAKERV-----FSFR 154  
 QY 155 RYLDQFVPEKMRRELKAYCGFLRPGVSSNLSAVATNCGGAFGGDARLALKALQILAA 214  
 Db 155 NAHRWDPNKORPELWNVNARIHPG-----ESRVFPLSNW----- 191  
 QY 215 AAAERDV-----VYFTGDSLMRDYISMHIFLTERKLTIV--GDVYKLL-LR 258  
 Db 192 --TELDVWLVIYREKIPVVPVLYFAAPRPVVERD--GMLILVDDERLPLRPGVSKLLWVR 247  
 QY 259 YNNEECRCNCSTPGPDIKLKYPIYIYNAVESCATAD 292  
 Db 248 FRTLGC-----YP-LTGAVESRAATLE 268  
 RESULT 9

YCF2\_LOTJA STANDARD; PRT; 2298 AA.  
 ID YCF2\_LOTJA  
 AC Q9B1K6;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Hypothetical 270.7 kDa protein ycf2.  
 GN YCF2-A AND YCF2-B.  
 OS Lotus japonicus.  
 OC Chloroplast.  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;  
 OC eurosids I; Fabales; Fabaceae; Papilionoideae; Lotaeae; Lotus.  
 OX NCBI\_TaxID=34305;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Accession MG-20;  
 MEDLINE=21082929; PubMed=11214967;  
 RX Kato T., Kaneko T., Sato S., Nakamura Y., Tabata S.;  
 RA "Complete structure of the chloroplast genome of a legume, Lotus  
 RT japonicus." RT  
 RL DNA Res. 7:323-330(2000).  
 CC -!- SIMILARITY: Belongs to the ycf2 family.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; AP002983; BAB33238.1; -.  
 CC EMBL; AP002983; BAB33256.1; -.  
 CC InterPro; IPR003593; AAA\_ATPase.  
 CC InterPro; IPR003959; AAA\_ATPase\_cent.  
 CC InterPro; IPR008543; DUF825.  
 CC Pfam; PF00004; AAA; 1.  
 CC Pfam; PF05695; DUF825; 1.  
 CC SMART; SM00382; AAA; 1.  
 CC Chloroplast; Hypothetical protein.  
 CC SEQUENCE 2298 AA; 270688 MW; 36E9A64CA54DC7D0 CRC64;  
 Query Match 5.3%; Score 84; DB 1; Length 2298;  
 Best Local Similarity 21.0%; Pred. No. 53;  
 Matches 47; Conservative 38; Mismatches 71; Indels 68; Gaps 10;  
 QY 2 TLF-----CYFRVTEK-----KPTGLVTFTRQSLDFFPEWERCEKPLRL- 42  
 Db 670 TLFYENESFYRIIKKWQVTSQCNLDLDPKP-IVVFTSNINAEVQYELIENLPIQ 728  
 QY 43 HVTYEGTIEENGQMLQVDFANRFVGGVTSAGLVOEIRFLINPELLIISR--LFTVLD 100  
 Db 729 YITY-----GYGIR-NVLNRFI-----QKNRFDRNFYRIQRYQIENDTLN 769  
 QY 101 H-----NECLITGTEQYSEYTGVAETYRWSRSHEDGSDERD 136  
 Db 770 HRTMKYTIINQHSNLSKKSKQKWFDSLILISTERSMNRDPNARYIKWNGNKNFQHL 829  
 QY 137 D---CERRCTEIVADLHPRYLDQFVPEKMRRELKAYCGFL 177  
 Db 830 DHRTSEQNSRFQVDFELRNQYSIDWSEVIDKKLSKSLCFFL 873  
 RESULT 10  
 RZPD\_ECOLI STANDARD; PRT; 153 AA.  
 ID RZPD\_ECOLI  
 AC P75719; P77084;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Putative R2 endopeptidase from lambdaoid prophage DLP12 (EC 3.4.-.-).  
 GN RZPD OR B0556.



```

Query Match      5.2%; Score 82.5; DB 1; Length 624;
Best Local Similarity 21.9%; Pred. No. 14;
Matches 43; Conservative 25; Mismatches 71; Indels 57; Gaps 10;

Qy   5 CYFRRVTEKPTGLVTFTQSLEDFEWEWCERCEPLTLRLHYTSG---TIENGOGMLQVD 61
     |||::||::||::||::|
Db   43 CYRSLVDTN-----YIFNK---DDF-----YIRISADKENKITLITDTGLGWTXDD 86

Qy   62 FANRFVGGVTSAGLVQEHRIFLINPELLIISRLPTEVLHDNECLITGTQSYSEYTYG-- 119
     ::||::|::|::|::|::|
Db   87 LENNL-GTTAKSG-----SFAFKSENAKEGVDIIGQFGVGFYSAPMV 128

Qy   120 -----AETVRSRSHEDGSERDDCRRC--TRIV-AIDALHFRRYLDQFVP 162
       ::||::|::|::|::|::|
Db   129 ADDVTWISRSVDSEEAAYKESKGVEGYTHCKBKTPTGTVILVKIKENTDDEKY-DEFELD 187

Qy   163 EKMRRRELNKA YCGFLR 178
       ::||::|::|
Db   188 EYKIRSLIKKYSDFIK 203

```

```

RESULT 13
AKA3 BOVIN
ID AKA3_BOVIN STANDARD; PET; 854 AA.
AC O77797;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE A-kinase anchor protein 3 (Protein kinase A anchoring protein 3)
DE (PRK3) (A-kinase anchor protein 110 kDa) (AKAP 110) (Fragment).
GN AKAP3 OR AKAP110.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Rutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
SEQUENCE FROM N.A.
RC TISSUE=Testis;
RX MEDLINE=99252879; PubMed=10319321;
RA Vijayaraghavan S., Liberty G.A., Mohan J., Winfrey V.P., Olson G.E.,
RA Carr D.W.;
RT "Isolation and molecular characterization of AKAP110, a novel,
RT sperm-specific protein kinase A-anchoring protein.";
RL Mol. Endocrinol. 13:705-717(1999).
CC -1- FUNCTION: May function as a regulator of both motility- and head-
CC associated functions such as capacitation and the acrosome
CC reaction (By similarity).
CC -1- SUBCELLULAR LOCATION: Localizes to the ribs of the fibrous sheath
CC in the principal piece of the sperm tail (By similarity).
CC -1- DOMAIN: RII-binding site, predicted to form an amphipathic
CC helix, could participate in protein-protein interactions with a
CC complementary surface on the R-subunit dimer.
CC -1- PTM: Phosphorylated on tyrosine (By similarity).
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF093407; AAC63370.1; -.
CC InterPro; IPR008382; AKAP_110.
CC Pfam; PF05716; AKAP_110; 1.
CC Phosphorylation.
CC NON TER 1
CC DOMAIN 127 140 PKA-RII SUBUNIT BINDING DOMAIN.
CC SEQUENCE 854 AA; 94216 MW; 0AB67E6E0327A32B CRC64;
CC -----
Query Match 5.2%; Score 82.5; DB 1; Length 854;
Best Local Similarity 23.0%; Pred. No. 20;

```





GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 26, 2004, 18:18:25 ; Search time 23.942 Seconds  
(without alignments)  
3953.539 Million cell updates/sec

Title: US-09-302-812-4\_COPY\_677\_977  
Perfect score: 1597  
Sequence: 1 KTLFCYFRRVTEKKPTGLVT.....YHAVESCAETADHSQRTGT 300

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues  
Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SPTREMBL 25:\*

- 1: sp\_archaea.\*
- 2: sp\_bacteria.\*
- 3: sp\_fungi.\*
- 4: sp\_human.\*
- 5: sp\_invertebrate.\*
- 6: sp\_mammal.\*
- 7: sp\_mhc.\*
- 8: sp\_organelle.\*
- 9: sp\_phase.\*
- 10: sp\_plant.\*
- 11: sp\_rodent.\*
- 12: sp\_virus.\*
- 13: sp\_vertebrate.\*
- 14: sp\_unclassified.\*
- 15: sp\_rvirus.\*
- 16: sp\_bacteriap.\*
- 17: sp\_archaea.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1597	100.0	976	4 Q9Y4W7	Q9Y4W7 homo sapien
2	1583	99.1	976	4 Q86W56	Q86W56 homo sapien
3	1578	98.8	976	4 Q7Z742	Q7Z742 homo sapien
4	1497	93.7	968	11 Q86G22	Q86G22 mus musculus
5	1493	93.5	977	6 O02776	O02776 bos taurus
6	1476	92.4	961	11 Q80YQ6	Q80YQ6 mus musculus
7	1464	91.7	972	11 Q9QYM2	Q9QYM2 rattus norv
8	1260	78.9	920	11 Q8C872	Q8C872 mus musculus
9	619.5	38.8	723	5 Q960N8	Q960N8 drosophila
10	619.5	38.8	768	5 Q46043	Q46043 drosophila
11	470	29.4	548	10 Q9SKB3	Q9SKB3 arabidopsis
12	429	26.9	522	10 Q8VYAL	Q8VYAL arabidopsis
13	390.5	24.5	485	5 Q9N5L4	Q9N5L4 caenorhabdi
14	381.5	23.9	764	5 Q19637	Q19637 caenorhabdi
15	381.5	23.9	781	5 Q867X0	Q867X0 caenorhabdi
16	320.5	20.1	368	5 Q86G14	Q86G14 toxoplasma

17	254.5	15.9	364	10 Q9SKB4	Q9SKB4 arabidopsis
18	131.5	8.2	508	12 Q9E234	Q9E234 helicoverpa
19	127.5	8.0	510	12 Q99G09	Q99G09 helicoverpa
20	103	6.4	579	10 Q94FW3	Q94FW3 abies grand
21	97.5	6.1	395	16 Q7VWX5	Q7VWX5 bordetella
22	97.5	6.1	401	16 Q7WH46	Q7WH46 bordetella
23	97.5	6.1	401	16 Q7W9L2	Q7W9L2 bordetella
24	96	6.0	955	16 Q89VU0	Q89VU0 bradyrhizob
25	95.5	6.0	2498	12 Q86924	Q86924 aura virus
26	94.5	5.9	756	16 Q82TY2	Q82TY2 nitrosomona
27	92.5	5.8	7349	2 Q8G6P2	Q8G6P2 streptomyces
28	90	5.6	602	12 Q85169	Q85169 african swi
29	90	5.6	1194	5 Q27826	Q27826 euplotes oc
30	89	5.6	543	16 Q8RC12	Q8RC12 thermoanaer
31	89	5.6	614	12 Q8V9S9	Q8V9S9 african swi
32	89	5.6	712	5 Q8NDI2	Q8NDI2 babesia bov
33	88	5.6	770	10 Q7X6G4	Q7X6G4 oryza sativ
34	88.5	5.5	263	11 Q99M28	Q99M28 rattus norv
35	88.5	5.5	2135	5 Q61077	Q61077 plasmodium
36	88	5.5	268	16 Q7ULC5	Q7ULC5 rhodospirill
37	88	5.5	448	16 Q8UFR7	Q8UFR7 agrobacteri
38	87.5	5.5	391	16 Q8VPR3	Q8VPR3 anabaena sp
39	87.5	5.5	541	16 Q7UJL3	Q7UJL3 rhodospirill
40	86.5	5.4	153	16 Q8X704	Q8X704 escherichia
41	86.5	5.4	435	16 Q89SS4	Q89SS4 escherichia
42	86.5	5.4	915	5 Q26872	Q26872 bradyrhizob
43	86	5.4	157	16 Q8FI80	Q8FI80 trypanosoma
44	86	5.4	510	10 Q9LJY7	Q9LJY7 escherichia
45	85.5	5.4	343	17 Q8TVQ9	Q8TVQ9 arabidopsis
					Q8TVQ9 methanopyru

ALIGNMENTS

RESULT 1

Q9Y4W7 PRELIMINARY; PRT; 976 AA.  
AC Q9Y4W7  
DT 01-NOV-1999 (TrEMBLrel. 12, Created)  
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
DE Poly(ADP-ribose) glycohydrolase.  
GN HPARG.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=99380098; PubMed=10449915;  
RA Ame J.C., Apolou F., Jacobson E.L., Jacobson M.K.;  
RT "Assignment of the poly(ADP-ribose) glycohydrolase gene (PARG) to  
human chromosome 10q11.23 and mouse chromosome 14B by in situ  
hybridization.";  
RL Cytogenet. Cell Genet. 85:269-270(1999).  
DR EMBL; AF005043; AAB61614.1; -;  
DR Genbank; HGNC:8605; PARG.  
DR GO; GO:0005737; C:cytoplasm; TAS.  
DR GO; GO:0004649; F:poly(ADP-ribose) glycohydrolase activity; TAS.  
DR InterPro; IPR007724; PARG.  
DR Pfam; PF05028; PARG; 1.  
KW Hydrolase.  
SQ SEQUENCE 976 AA; 110903 MW; 42A0F51466FCE7DC CRC64;

Query Match 100.0%; Score 1597; DB 4; Length 976;  
Best Local Similarity 100.0%; Pred. No. 1.4e-144;  
Matches 300; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 KTLFCYFRRVTEKKPTGLVTFTQSLDFPEWERCERPLRLHVTYEGTTEENQGMLOV 60  
Db 677 KTLFCYFRRVTEKKPTGLVTFTQSLDFPEWERCERPLRLHVTYEGTTEENQGMLOV 736  
Qy 61 DFANRFGVGGVTSAGLVQEEIRFLINPELISRLFTFVLDHNECLLIITGTQYSEYTGVA 120

Db 737 DFANFVGGVTSAGLVQEEIRFLINPELIIISRLFTEVLDDHNECLIIITGEQYSEYTGVA 796  
QY 121 ETYRWSRSHEDGSRDDCERRCTEIVADALHFRYLDQFVPEKMRRELNKAYCGFLRPG 180  
Db 797 ETYRWSRSHEDGSRDDCERRCTEIVADALHFRYLDQFVPEKMRRELNKAYCGFLRPG 856  
QY 181 VSSENLSAVATGNWCGGAFGGDARLKALIIILAAAAAERDVVYFTFGDSELMRDYISMH 240  
Db 857 VSSENLSAVATGNWCGGAFGGDARLKALIIILAAAAAERDVVYFTFGDSELMRDYISMH 916  
QY 241 FLTERKLTVDGVYKLLRYNEECNCSCTPGPDIKLYPIYHVESCAETADHSQRTGT 300  
Db 917 FLTERKLTVDGVYKLLRYNEECNCSCTPGPDIKLYPIYHVESCAETADHSQRTGT 976

RESULT 2  
Q86W56 PRELIMINARY; PRT; 976 AA.  
AC Q86W56;  
DT 01-JUN-2003 (TREMELrel. 24, Created)  
DT 01-JUN-2003 (TREMELrel. 24, Last sequence update)  
DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)  
DE Similar to poly (ADP-ribose) glycohydrolase.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Skin;  
RA Strausberg R.;  
RL Submitted (APR-2003) to the EMBL/GenBank/DBJ databases.  
DR EMBL; BC050560; AAH50560.1; -  
DR GO; GO:0016787; F:hydrolase activity; IEA.  
DR InterPro; IPR007724; PARG.  
DR Pfam; PF05028; PARG; 1.  
KW Hydrolase.  
SQ SEQUENCE 976 AA; 111109 MW; D6646353C6D0180E CRC64;

Query Match 99.1%; Score 1583; DB 4; Length 976;  
Best Local Similarity 99.3%; Pred. No. 3.1e-143;  
Matches 298; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
QY 1 KTLFCYFRVTEKKPTGLVTFTQSLDFPWEKCEKPLTRLHVYEGTIENGQMLQV 60  
Db 677 KTLFCYFRVTEKKPTGLVTFTQSLDFPWEKCEKPLTRLHVYEGTIENGQMLQV 736  
QY 61 DFANFVGGVTSAGLVQEEIRFLINPELIIISRLFTEVLDDHNECLIIITGEQYSEYTGVA 120  
Db 737 DFANFVGGVTSAGLVQEEIRFLINPELIIISRLFTEVLDDHNECLIIITGEQYSEYTGVA 796  
QY 121 ETYRWSRSHEDGSRDDCERRCTEIVADALHFRYLDQFVPEKMRRELNKAYCGFLRPG 180  
Db 797 ETYRWSRSHEDGSRDDCERRCTEIVADALHFRYLDQFVPEKMRRELNKAYCGFLRPG 856  
QY 181 VSSENLSAVATGNWCGGAFGGDARLKALIIILAAAAAERDVVYFTFGDSELMRDYISMH 240  
Db 857 VSSENLSAVATGNWCGGAFGGDARLKALIIILAAAAAERDVVYFTFGDSELMRDYISMH 916  
QY 241 FLTERKLTVDGVYKLLRYNEECNCSCTPGPDIKLYPIYHVESCAETADHSQRTGT 300  
Db 917 FLTERKLTVDGVYKLLRYNEECNCSCTPGPDIKLYPIYHVESCAETADHSQRTGT 976

RESULT 3  
Q72742 PRELIMINARY; PRT; 976 AA.  
ID Q72742  
AC Q72742;  
DT 01-OCT-2003 (TREMELrel. 25, Created)  
DT 01-OCT-2003 (TREMELrel. 25, Last sequence update)  
DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)  
DE Hypothetical protein.

OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Testis;  
RA MEDLINE=22389257; PubMed=12477932;  
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
RA Hopkins R.F., Moore T., Max S.I., Wang J., Hsieh F.,  
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.B.,  
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,  
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,  
RA Bosak S.A., McSwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
RA Villalon D.K., Muzny K.C., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Fahy J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,  
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butlerfield Y.S.,  
RA Krzywinski M.I., Skalska U., Smalusz D.E., Schnerch A., Schein J.E.,  
RA Jones S.J., Marra M.A.;  
RT "Generation and initial analysis of more than 15,000 full-length human  
ET and mouse cDNA sequences";  
FL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Testis;  
RA Strausberg R.;  
RL Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.  
DR EMBL; BC052966; AAH52966.1; -  
KW Hypothetical protein.  
SQ SEQUENCE 976 AA; 111090 MW; 47C8EE1826C4A74A CRC64;

Query Match 98.8%; Score 1578; DB 4; Length 976;  
Best Local Similarity 99.0%; Pred. No. 9.4e-143;  
Matches 297; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
QY 1 KTLFCYFRVTEKKPTGLVTFTQSLDFPWEKCEKPLTRLHVYEGTIENGQMLQV 60  
Db 677 KTLFCYFRVTEKKPTGLVTFTQSLDFPWEKCEKPLTRLHVYEGTIENGQMLQV 736  
QY 61 DFANFVGGVTSAGLVQEEIRFLINPELIIISRLFTEVLDDHNECLIIITGEQYSEYTGVA 120  
Db 737 DFANFVGGVTSAGLVQEEIRFLINPELIIISRLFTEVLDDHNECLIIITGEQYSEYTGVA 796  
QY 121 ETYRWSRSHEDGSRDDCERRCTEIVADALHFRYLDQFVPEKMRRELNKAYCGFLRPG 180  
Db 797 ETYRWSRSHEDGSRDDCERRCTEIVADALHFRYLDQFVPEKMRRELNKAYCGFLRPG 856  
QY 181 VSSENLSAVATGNWCGGAFGGDARLKALIIILAAAAAERDVVYFTFGDSELMRDYISMH 240  
Db 857 VSSENLSAVATGNWCGGAFGGDARLKALIIILAAAAAERDVVYFTFGDSELMRDYISMH 916  
QY 241 FLTERKLTVDGVYKLLRYNEECNCSCTPGPDIKLYPIYHVESCAETADHSQRTGT 300  
Db 917 FLTERKLTVDGVYKLLRYNEECNCSCTPGPDIKLYPIYHVESCAETADHSQRTGT 976

RESULT 4  
O88622 PRELIMINARY; PRT; 968 AA.  
ID O88622  
AC O88622;  
DT 01-NOV-1998 (TREMELrel. 08, Created)  
DT 01-NOV-1998 (TREMELrel. 08, Last sequence update)  
DT 01-JUN-2003 (TREMELrel. 24, Last annotation update)  
DE Poly (ADP-ribose) glycohydrolase.  
GN PARG.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=99380098; PubMed=10449915;  
 RA Ame J.C., Apicou F., Jacobson E.L., Jacobson M.K.;  
 RT "Assignment of the poly(ADP-ribose) glycohydrolase gene (PARG) to  
 human chromosome 10q11.23 and mouse chromosome 14B by in situ  
 hybridization.";  
 RL Cytogenet. Cell Genet. 85:269-270(1999).  
 DR EMBL; AF079557; AAC28735.1; -;  
 DR MGI; MGI:1347094; PARG.  
 DR GO; GO:0016787; F:hydrolase activity; IEA.  
 DR InterPro; IPR007724; PARG.  
 DR Pfam; PF05028; PARG; 1.  
 KW Hydrolase.  
 SQ SEQUENCE 968 AA; 109169 MW; 3981DA4809E1E007 CRC64;

Query Match 93.7%; Score 1497; DB 11; Length 968;  
 Best Local Similarity 93.7%; Pred. No. 5.9e-135;  
 Matches 281; Conservative 7; Mismatches 12; Indels 0; Gaps 0;

QY 1 KTLFCYFRVTEKKPTGLVTFTRQSLDEPPEWERCCKPLTRLHVTVYEGTIEGNGQMLQV 60  
 Db 669 KTLFCYFRVTEKKPTGLVTFTRQSLDEPPEWERCCKPLTRLHVTVYEGTIEGNGQMLQV 728

QY 61 DFANRFVGGVTSAGLVQVEIRFLINPELIIISRLFTVLDHNECLIITGTEQYSEYTGVA 120  
 Db 729 DFANRFVGGVTSAGLVQVEIRFLINPELIIISRLFTVLDHNECLIITGTEQYSEYTGVA 788

QY 121 EYWRSHEDGSDRDCERRCTEIVADALHFRYLDQFVPEKMRRELKAYCGFLRPG 180  
 Db 789 EYWRSHEDGSEKDDQWRCTEIVADALHFRYLDQFVPEKMRRELKAYCGFLRPG 848

QY 181 VSENLSAVATGNWCGAFGGDARLKALIQILAAAAARDVYVFTFGDSELMRDIIYSMT 240  
 Db 849 VSENLSAVATGNWCGAFGGDARLKALIQILAAAAARDVYVFTFGDSELMRDIIYSMT 908

QY 241 FLTERKLTVDVYKLLRYNNEECRCNSTPGDPDKLYPFIYHAVESCAETADHSGQRTGT 300  
 Db 909 FLTERKLDGVKYLKLLRYNNEECRCNSTPGDPDKLYPFIYHAVESCAETADHSGQRTGT 968

RESULT 5  
 O02776 PRELIMINARY; PRT; 977 AA.  
 AC O02776;  
 DT 01-JUL-1997 (TrEMBLrel. 04, Created)  
 DT 01-JUL-1997 (TrEMBLrel. 04, Last sequence update)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
 DE Poly(ADP-ribose) glycohydrolase.  
 GN BPARG.  
 OS Bos taurus (Bovine).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;  
 OC Bovidae; Bovinae; Bos.  
 OX NCBI\_TaxID=9913;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=9277329; PubMed=9115250;  
 RA Lin W., Ame J.C., Aboul-Ela N., Jacobson E.L., Jacobson M.K.;  
 RT "Isolation and characterization of the cDNA encoding bovine poly(ADP-ribose) glycohydrolase.";  
 RL J. Biol. Chem. 272:11895-11901(1997).  
 DR EMBL; U78975; AAB53370.1; -;  
 DR GO; GO:0016787; F:hydrolase activity; IEA.  
 DR InterPro; IPR007724; PARG.  
 DR Pfam; PF05028; PARG; 1.  
 KW Hydrolase.  
 SQ SEQUENCE 977 AA; 110837 MW; 87D2100F979DF377 CRC64;

Query Match 93.5%; Score 1493; DB 6; Length 977;  
 Best Local Similarity 94.0%; Pred. No. 1.4e-134;

Matches 281; Conservative 7; Mismatches 11; Indels 0; Gaps 0;

QY 1 KTLFCYFRVTEKKPTGLVTFTRQSLDEPPEWERCCKPLTRLHVTVYEGTIEGNGQMLQV 60  
 Db 678 KTLFCYFRVTEKKPTGLVTFTRQSLDEPPEWERCCKPLTRLHVTVYEGTIEGNGQMLQV 737

QY 61 DFANRFVGGVTSAGLVQVEIRFLINPELIIISRLFTVLDHNECLIITGTEQYSEYTGVA 120  
 Db 738 DFANRFVGGVTSAGLVQVEIRFLINPELIIISRLFTVLDHNECLIITGTEQYSEYTGVA 797

QY 121 EYWRSHEDGSDRDCERRCTEIVADALHFRYLDQFVPEKMRRELKAYCGFLRPG 180  
 Db 798 EYWRSHEDGSDRDCERRCTEIVADALHFRYLDQFVPEKMRRELKAYCGFLRPG 857

QY 181 VSENLSAVATGNWCGAFGGDARLKALIQILAAAAARDVYVFTFGDSELMRDIIYSMT 240  
 Db 858 VSENLSAVATGNWCGAFGGDARLKALIQILAAAAARDVYVFTFGDSELMRDIIYSMT 917

QY 241 FLTERKLTVDVYKLLRYNNEECRCNSTPGDPDKLYPFIYHAVESCAETADHSGQRTGT 299  
 Db 918 FLTERKLTVDVYKLLRYNNEECRCNSTPGDPDKLYPFIYHAVESCAETADHSGQRTGT 976

RESULT 6  
 Q80YQ6 PRELIMINARY; PRT; 961 AA.  
 ID Q80YQ6;  
 DT 01-JUN-2003 (TrEMBLrel. 24, Created)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Poly (ADP-ribose) glycohydrolase.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6; TISSUE=Brain;  
 RA Strausberg R.;  
 RL Submitted (APR-2003) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; BC050892; AAB50892.1; -;  
 DR GO; GO:0016787; F:hydrolase activity; IEA.  
 DR InterPro; IPR007724; PARG.  
 DR Pfam; PF05028; PARG; 1.  
 KW Hydrolase.  
 SQ SEQUENCE 961 AA; 108580 MW; 1558F120C367CF63 CRC64;

Query Match 92.4%; Score 1476; DB 11; Length 961;  
 Best Local Similarity 94.9%; Pred. No. 6.1e-133;  
 Matches 277; Conservative 6; Mismatches 9; Indels 0; Gaps 0;

QY 1 KTLFCYFRVTEKKPTGLVTFTRQSLDEPPEWERCCKPLTRLHVTVYEGTIEGNGQMLQV 60  
 Db 670 KTLFCYFRVTEKKPTGLVTFTRQSLDEPPEWERCCKPLTRLHVTVYEGTIEGNGQMLQV 729

QY 61 DFANRFVGGVTSAGLVQVEIRFLINPELIIISRLFTVLDHNECLIITGTEQYSEYTGVA 120  
 Db 730 DFANRFVGGVTSAGLVQVEIRFLINPELIIISRLFTVLDHNECLIITGTEQYSEYTGVA 789

QY 121 EYWRSHEDGSDRDCERRCTEIVADALHFRYLDQFVPEKMRRELKAYCGFLRPG 180  
 Db 790 EYWRSHEDGSEKDDQWRCTEIVADALHFRYLDQFVPEKMRRELKAYCGFLRPG 849

QY 181 VSENLSAVATGNWCGAFGGDARLKALIQILAAAAARDVYVFTFGDSELMRDIIYSMT 240  
 Db 850 VSENLSAVATGNWCGAFGGDARLKALIQILAAAAARDVYVFTFGDSELMRDIIYSMT 909

QY 241 FLTERKLTVDVYKLLRYNNEECRCNSTPGDPDKLYPFIYHAVESCAETAD 292  
 Db 910 FLTERKLDGVKYLKLLRYNNEECRCNSTPGDPDKLYPFIYHAVESCAETAD 961

RESULT 7



Q9QYM2 PRELIMINARY; PRT; 972 AA.

AC Q9QYM2; 01-MAY-2000 (TREMELrel. 13, Created)

DT 01-MAY-2000 (TREMELrel. 13, Last sequence update)

DT 01-JUN-2003 (TREMELrel. 24, Last annotation update)

DE Poly (ADP-ribose) glycohydrolase.

GN PARG.

OS Rattus norvegicus (Rat).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.

OX NCBI\_TaxID=10116;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=BUF; TISSUE=Colon;

RA Shimokawa T., Maebuchi M., Nagasawa S., Nozaki T., Ikota N., Araki S.,

RA Aoki Y., Nakagawa H., Sugimura T.;

RT "Isolation and cloning of rat poly (ADP-ribose) glycohydrolase

RT (Parg).";

RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.

DR EMBL; AB019366; BAA87901.1; --

DR GO; GO:0016787; F:hydrolase activity; IEA.

DR InterPro; IPR007724; PARG.

DR Pfam; PF05028; PARG; 1.

KW Hydrolase.

QY SEQUENCE 972 AA; 109002 MW; 3DED7885684F59E9 CRC64;

Query Match 91.7%; Score 1464; DB 11; Length 972;

Best Local Similarity 91.0%; Pred. No. 8.9e-132;

Matches 273; Conservative 12; Mismatches 15; Indels 0; Gaps 0;

QY 1 KTLFCYFRVTEKPTGLVTFTRQSLDEDFPWERCEKPLTRHLVYEGTIRENGGMLQV 60

DB 673 KTLFCYFRVTEKPTGLVTFTRQSLDEDFPWERCEKPLTRHLVYEGTIRENGGMLQV 732

QY 61 DPANFVGGVTSAGLVQEEIRFLINPELIIISRLFTFVLDHNECLIIITGTEYSYTGVA 120

DB 733 DPANFVGGVTSAGLVQEEIRFLINPELIIISRLFTFVLDHNECLIIITGTEYSYTGVA 792

QY 121 ETYRWSRSHEDGSDRDCRCECTEIVADLHFRYLDQFVPEKVRRELINKAYCGFLRPG 180

DB 793 ETYRWSRSHEDGSDRDCRCECTEIVADLHFRYLDQFVPEKVRRELINKAYCGFLRPG 852

QY 181 VSENLSAVATGNWCGAGFGDARLKALIQILAAAAERDVVYFTFGSELMDIYSMEI 240

DB 853 VPPENLSAVATGNWCGAGFGDARLKALIQILAAAAERDVVYFTFGSELMDIYSMEI 912

QY 241 FLTERKLTVDGVKLLRYNBEKNCSTPGPDIKLYPIYHVSACPTADHSQORTGT 300

DB 913 FLTERKLTVDGVKLLRYNBEKNCSTPGPDIKLYPIYHVSACPTADHSQORTGT 972

RESULT 8

Q8CB72 PRELIMINARY; PRT; 920 AA.

AC Q8CB72; 01-MAR-2003 (TREMELrel. 23, Created)

DT 01-MAR-2003 (TREMELrel. 23, Last sequence update)

DT 01-JUN-2003 (TREMELrel. 24, Last annotation update)

DE Poly.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI\_TaxID=10090;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=C57BL/6J; TISSUE=Bone;

RA MEDLINE=22354683; PubMed=12466851;

RA The FANTOM Consortium,

RA The RIKEN Genome Exploration Research Group Phase I & II Team;

RT "Analysis of the mouse transcriptome based on functional annotation of

RT 60,770 full-length cDNAs.";

RL Nature 420:563-573 (2002).

DR EMBL; AK036656; BAC29519.1; --

DR MGD; MGI:1347094; Parg.

DR InterPro; IPR007724; PARG.

DR Pfam; PF05028; PARG; 1.

SQ SEQUENCE 920 AA; 103799 MW; 379E779E121B09A1 CRC64;

Query Match 78.9%; Score 1260; DB 11; Length 920;

Best Local Similarity 94.4%; Pred. No. 3.7e-112;

Matches 237; Conservative 7; Mismatches 7; Indels 0; Gaps 0;

QY 1 KTLFCYFRVTEKPTGLVTFTRQSLDEDFPWERCEKPLTRHLVYEGTIRENGGMLQV 60

DB 670 KTLFCYFRVTEKPTGLVTFTRQSLDEDFPWERCEKPLTRHLVYEGTIRENGGMLQV 729

QY 61 DPANFVGGVTSAGLVQEEIRFLINPELIIISRLFTFVLDHNECLIIITGTEYSYTGVA 120

DB 730 DPANFVGGVTSAGLVQEEIRFLINPELIIISRLFTFVLDHNECLIIITGTEYSYTGVA 789

QY 121 ETYRWSRSHEDGSDRDCRCECTEIVADLHFRYLDQFVPEKVRRELINKAYCGFLRPG 180

DB 790 ETYRWSRSHEDGSDRDCRCECTEIVADLHFRYLDQFVPEKVRRELINKAYCGFLRPG 849

QY 181 VSENLSAVATGNWCGAGFGDARLKALIQILAAAAERDVVYFTFGSELMDIYSMEI 240

DB 850 VPPENLSAVATGNWCGAGFGDARLKALIQILAAAAERDVVYFTFGSELMDIYSMEI 909

QY 241 FLTERKLTVDGV 251

DB 910 FLTERKLTVDGV 920

RESULT 9

Q960N8 PRELIMINARY; PRT; 723 AA.

ID Q960N8

AC Q960N8; 01-DEC-2001 (TREMELrel. 19, Created)

DT 01-DEC-2001 (TREMELrel. 19, Last sequence update)

DT 01-JUN-2003 (TREMELrel. 24, Last annotation update)

DE LD4238Op.

GN PARG OR EG:114E2.1 OR CG2864.

OS Drosophila melanogaster (Fruit fly).

OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

OC Ephydroidea; Drosophilidae; Drosophila.

OX NCBI\_TaxID=7227;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=Berkley;

RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,

RA Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,

RA Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,

RA Nunoo J., Paclieb J., Paragas V., Park S., Phouanavong S., Wan K.,

RA Yu C., Lewis S.E., Rubin G.M., Celnikier S.;

RA Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.

DR EMBL; AY051955; AAK93379.1; --

DR FlyBase; FBgn0023216; Parg.

DR InterPro; IPR007724; PARG.

DR Pfam; PF05028; PARG; 1.

SQ SEQUENCE 723 AA; 81109 MW; 90A5BA59775D611C CRC64;

Query Match 38.8%; Score 619.5; DB 5; Length 723;

Best Local Similarity 46.4%; Pred. No. 1.3e-50;

Matches 140; Conservative 41; Mismatches 90; Indels 31; Gaps 8;

QY 1 KTLFCYFRV--TEKK-----FTGLVTFTRQS--LEDFFEWERCXPL--TRLHVTYEGTI 50

DB 252 KCIHMYFRVCPTEKASNVPTGVVTFVRSGLPEHLIDWSQSAAPLGDVPLHVDAGETI 311

QY 51 BENGOGMLQVDFANFVGGVTSAGLVQEEIRFLINPELIIISRLFTFVLDHNECLIIITGT 110

DB 312 EDEGILLQVDFANKYLGGLGVGHGCVQEEIRFVCPSELVGLKFTFELRPFALVMLGA 371

```

QY 111 EQSEYTGATYATWSRSHSDGSRDCCERCTEIVADALHFRYLDQFVPERRELN 170
D 372 ERYSNYTGAGSFWGNSFSTPRDSSGRQTAIVADALHFAQSHHGYEDLMERELN 431
QY 171 KAYCGFLR-----PGVSSNLSAVATGNCWCGAFGGDARLKALIQILAAAAERDVYF 224
D 432 KAYIGFVHMVTPPPG-----VATGNCWCGAFGGDSYLKALLQVMVCAQLGRPLAYY 483
QY 225 TFGDSELMRIYSMHIFLTERKLTWGDVYKLLRYNEECNCSTPGP-----DILKYP 278
D 484 TFGNVEFRDDFHEWMLLFRNDGTIVQQLWS--ILRSYSRLIKEKSKPRENKASKKLYD 542
QY 279 FI 280
D 543 FI 544

RESULT 10
O46043 PRELIMINARY; PRT; 768 AA.
ID O46043
AC O46043
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE PARG protein.
GN PARG OR EG:114E2.1 OR CG2864.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
[1]
RN SEQUENCE FROM N.A.
RC STRAIN=Berkely;
RX MEDLINE=2019606; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agabani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballwe R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Folsler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacieb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirkas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195 (2000).

```

```

RN SEQUENCE FROM N.A.
RA Ame J.-C., Jacobson M.K.;
RT "Isolation and characterization of the cDNA encoding Drosophila
RL poly(ADP-ribose) glycohydrolase."
RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
[3]
RN SEQUENCE OF 46-768 FROM N.A.
RA Murphy L., Harris D., Barrell B.;
RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.
[4]
RN SEQUENCE OF 46-768 FROM N.A.
RA Benos P.;
RL Submitted (DEC-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF003428; AAF45866.1; -
DR EMBL; AF079556; AAC28734.1; -
DR EMBL; Z98254; CAB10913.1; -
DR FlyBase; FBgn0023216; Parg.
DR GO; GO:0016787; F:hydrolase activity; IEA.
DR InterPro; IPR007724; PARG.
DR Pfam; PF05028; PARG; 1.
KW Hydrolase.
SQ SEQUENCE 768 AA; 86288 MW; 2E620C42E643FC04 CRC64;
Query Match 38.8%; Score 619.5; DB 5; Length 768;
Best Local Similarity 46.4%; Pred. No. 1.4e-50;
Matches 140; Conservative 41; Mismatches 90; Indels 31; Gaps 8;
QY 1 KTLFCFRRV--TEKK-----PTGLVTFTRQS--LEDFPEWERCCKPL--TRLHVTYEGTI 50
D 297 KCIHYFRRCVPTERDASNVFTGVTFVRRSGLPEHLIDWSQSAAPLDGVLPHVDABGTI 356
QY 51 EENGQGMQVDFANRFVGGVTSAGLVOBEIRFLINPELIIISRLFTVLVDHNECLITGT 110
D 357 EDEIGILLQVDFANKYLGGLGVCGVQBEIRFVICPELVKGLFTECLRPFFALVMLGA 416
QY 111 EQSEYTGATYATWSRSHSDGSRDCCERCTEIVADALHFRYLDQFVPERRELN 170
D 417 ERYSNYTGAGSFWGNSFSTPRDSSGRQTAIVADALHFAQSHHGYEDLMERELN 476
QY 171 KAYCGFLR-----PGVSSNLSAVATGNCWCGAFGGDARLKALIQILAAAAERDVYF 224
D 477 KAYIGFVHMVTPPPG-----VATGNCWCGAFGGDSYLKALLQVMVCAQLGRPLAYY 528
QY 225 TFGDSELMRIYSMHIFLTERKLTWGDVYKLLRYNEECNCSTPGP-----DILKYP 278
D 529 TFGNVEFRDDFHEWMLLFRNDGTIVQQLWS--ILRSYSRLIKEKSKPRENKASKKLYD 587
QY 279 FI 280
D 588 FI 589

RESULT 11
Q9SKB3 PRELIMINARY; PRT; 548 AA.
ID Q9SKB3; Q94ET7;
AC Q9SKB3; Q94ET7;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DE Putative poly(ADP-ribose) glycohydrolase.
GN AT2G31870 OR TEPJ.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsi.
OX NCBI_TaxID=3702;
[1]
RN SEQUENCE FROM N.A.
RC STRAIN=ev. Columbia;
RA Lin X., Kaul S., Shea T.P., Fujii C.Y., Shen M., VanAken S.E.,
RA Barnstead M.E., Mason T.M., Bowman C.L., Ronning C.M., Benito M.-I.,
RA Carrera A.J., Creasy T.H., Buell C.R., Town C.D., Nierman W.C.,

```

RA Fraser C.M., Venter J.C.;  
RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Panda S., Poirier G.G., Kay S.A.;  
RT "TEJ defines a role for poly-ADP-ribosylation in establishing period  
length of the Arabidopsis circadian oscillator."  
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AC006533; AAD32285.2; --  
DR EMBL; AF394690; AAK72256.1; --  
DR PIR; B84726; B84726.  
DR GO; GO:0016787; F:hydrolase activity; IEA.  
DR InterPro; IPR007724; PARG.  
DR Pfam; PF05028; PARG; 1.  
KW Hydrolase.  
SQ SEQUENCE 548 AA; 62169 MW; FIA79FDAL57C3229 CRC64;  
  
Query Match 29.4%; Score 470; DB 10; Length 548;  
Best Local Similarity 35.1%; Pred. No. 2.2e-36;  
Matches 115; Conservative 40; Mismatches 91; Indels 82; Gaps 7;  
  
QY 6 YFRVTEKPTGLVTFTRQ--SLDFPEWERCERKPLTELHYTGTIEENGQGMLOVDF 63  
DB 198 YFERPCSCVPIGIVSFERKITAADPFWKSLDVSCLCAFKVHSFGLIEDQPDNALEVDFA 257  
  
QY 64 NRVFVGGSAGLVQEEIRFLINPELIISRLTFEVLHDNECLITGTQYSEYTYGATY 123  
DB 258 NKYLGGSLSGCVQEEIRFMINPELIAGMLFLPRMDNEAIEIVGAERFSCYTYGASSF 317  
  
QY 124 RWSRSHEDGSRDRCRCTEIVAI--HFRYLDQFVPEKMRRLNKAQCGFL 177  
DB 318 RFAGEYIDKKAMPFKRRTRIVAI--LLREINKALCGFL 370  
  
QY 178 R----- 178  
DB 371 NCSKAMEHQNFMEQDNEIQLVNRGRDGLLRTTETASHHTPLNDVEMNEKPAANLIR 430  
  
QY 179 ----PGVSSENL--SAVATGNGGAGFGDARLKALIQILAAARLDVV-YFTFGDSEL 231  
DB 431 DFYVEGVNDEHDDGVATGNGGCVFGDPELKATIQWLAASTQRRPFISYITFG-VEA 489  
  
QY 232 MRDIYSMHIFLTERKLTVDVYKLLRY 259  
DB 490 LRNLDDQVTKWLSHKWTVGDLNMMWLEY 517  
  
RESULT 12  
Q8VYAL PRELIMINARY; PRT; 522 AA.  
ID Q8VYAL  
AC Q8VYAL;  
DT 01-MAR-2002 (TrEMBLrel. 20, Created)  
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
DE Putative poly(ADP-ribose) glycohydrolase.  
GN AT2G31870.  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;  
OC eucosids II; Brassicales; Brassicaceae; Arabidopsia.  
OX NCBI\_TaxID=3702;  
[1]  
RP SEQUENCE FROM N.A.  
RA Nguyen M., Karlin-Neumann G., Southwick A., Lam B., Miranda M.,  
RA Palm C.J., Bowser L., Jones T., Banh J., Carninci P., Chen H.,  
RA Cheuk R., Chung M.K., Hayaashizaki Y., Ishida J., Kamiya A., Kawai J.,  
RA Kim C., Lin J., Liu S.X., Narusaka M., Pham P.K., Sakano H.,  
RA Sakurai T., Satou M., Seki M., Shinn P., Yamada K., Shinozaki K.,  
RA Ecker J., Theologis A., Davis R.W.;  
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AY072330; AAL61937.1; --  
DR GO; GO:0016787; F:hydrolase activity; IEA.  
DR InterPro; IPR007724; PARG.  
DR Pfam; PF05028; PARG; 1.  
DR

KW Hydrolase.  
SQ SEQUENCE 522 AA; 58942 MW; 6FF98685D867CA15 CRC64;  
  
Query Match 26.9%; Score 429; DB 10; Length 522;  
Best Local Similarity 32.9%; Pred. No. 1.8e-32;  
Matches 102; Conservative 56; Mismatches 102; Indels 50; Gaps 7;  
  
QY 1 KTLFCFRRRYTEKPTGLVTFTRQSLF-----DFPE--WERCCKPLRLHVTVEGT 49  
DB 195 KCLHYFGRICRWMPITGTVSFERKILPLEYHPFVSYPKADSWANSVTPLCSIEHTSGA 254  
  
QY 50 IEENGQMLQVDFANRFVGGVTSAGLVQEEIRFLINPELIISRLTFEVLHDNECLITG 109  
DB 255 IEDQPCALEVDFAEYFGGLTSLDYTLQEEIRFVINPELIAGMIFLPRMDANEAIIVG 314  
  
QY 110 TEQYSEYTVGAETVRWSRSHEDGSRDRCRCTEIVAI--HFRYLDQFVPEKMRRL 169  
DB 315 VERFSGYTGPGSFQVAGDYTDNKLDIFRRKTRVIAIDAMP-DPGMGQYKLDALIREV 373  
  
QY 170 NKAYCGFLR-----PGVSS-----ENLSAVATGN 193  
DB 374 NKAFSGYMHQCKYNIDVKHDPKASSSHVPLTSDASQVIESSHRWCIHBEKKIGVATGN 433  
  
QY 194 WGGCAFCDARLKALIQILAAARLDVV-YFTFGDSELMDIYSMHIFLTERKLTVDV 252  
DB 434 WGGCVFGDPELKIMLQWLAIQSGRPFMSYITFG-LQALQNUNQVIMVALQENTVGD 492  
  
QY 253 YKLILRYNE 262  
DB 493 WKLVYEISSE 502  
  
RESULT 13  
Q9NSL4 PRELIMINARY; PRT; 485 AA.  
ID Q9NSL4  
AC Q9NSL4;  
DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
DE Hypothetical protein H23L24.5 (Poly ADP-ribose metabolism enzyme  
4).  
GN H23L24.5 OR PME-4.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
OC Rhabditidae; Peloderinae; Caenorhabditis.  
OX NCBI\_TaxID=6239;  
[1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Bristol N2;  
RX MEDLINE=99069613; PubMed=9851916;  
RA Waterston R.;  
RT "Genome sequence of the nematode C. elegans: a platform for  
investigating biology. The C. elegans Sequencing Consortium.";  
RL Science 282:2012-2018(1998).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Bristol N2;  
RA Abu-Threideh J., Lehnert I.;  
RT "The sequence of C. elegans cosmid H23L24.";  
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Bristol N2;  
RA Waterston R.;  
RL Submitted (DEC-2002) to the EMBL/GenBank/DBJ databases.  
RN [4]  
RP SEQUENCE FROM N.A.  
RA Gagnon S.N., Hardy I., Desnoyers S.;  
RT "Molecular characterization of poly ADP-ribose glycohydrolases in  
Caenorhabditis elegans.";  
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AC006662; AAF39896.2; --  
DR EMBL; AF548468; AAN40699.1; --



```
Db      591 YIGYCHTLKWAELQPNHSRQNTNEFDREGRLERVETIAIDAILFKGSKLDCQTEQLNKAN 650
QY      165 MRRELKAYCGFLRPGVSSSENLSAVATGNWGCAGCGDARLKALIQILAAAAAERDVVYF 224
Db      651 IIREMKASIGFMSQGPKEFTNI-PIVTGNWGCAGFNGDKPLKFIIQVIAAGVADRLHFC 709
QY      225 TRQDSELMRDIYSMHIFLTERKLTUGDVYKLLLRYYNEBCRNCSTPGPDIKLYPF 279
Db      710 SFGEPELAAKCKIIERMKQKQVTLGMLFSMI-----NNTGLPHKHFEFYVF 756
```

Search completed: May 26, 2004, 18:46:02  
Job time : 26.942 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: May 26, 2004, 15:28:37 ; Search time 162.413 Seconds

(without alignments)  
13270.296 Million cell updates/sec

Title:

Perfect score: 6773

Sequence: 1 gggggactgtgtcgtcgggg.....aatcattgtcagaaaaa 3814

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 3172214

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+,n2p,model -DEV=xlp  
-Q=/cgn2\_1/USPRO\_spool\_p/633148/runat\_26052004\_150052\_5654/app\_query.fasta.1.12437  
-DB=A\_Geneseq\_29Jan04 -Qfmt=faстан -SUFFIX=n2p.rag -MINMATCH=0.1 -LOOPEL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFWT=plp -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=200000000  
-USER=633148\_@CGN\_1\_1\_414\_@runat\_26052004\_150052\_5654 -NCPU=6 -ICPU=3  
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSFBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

A\_Geneseq\_29Jan04.\*  
1: Geneseqp1980s.\*  
2: Geneseqp1990s.\*  
3: Geneseqp2000s.\*  
4: Geneseqp2001s.\*  
5: Geneseqp2002s.\*  
6: Geneseqp2003as.\*  
7: Geneseqp2003bs.\*  
8: Geneseqp2004s.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	5164	76.2	968	5 AAE25631	Murine po
2	5164	76.2	968	5 AAU76022	Mouse pol
3	5164	76.2	968	5 AAU76013	Mouse pol
4	5164	76.2	968	6 ABG72280	Mmurine p
5	4395	64.9	976	5 AAE25630	Human pol
6	4395	64.9	976	5 AAU76021	Human pol
7	4395	64.9	976	5 AAU76012	Human pol
8	4395	64.9	976	6 ABG72279	Human pol
9	4332.5	64.0	977	5 AAE25629	Bovine po
10	4332.5	64.0	977	5 AAU76020	Bovine po

11	4332.5	64.0	977	5	AAU75799	Bovine po
12	4332.5	64.0	977	6	ABG72278	Bovine po
13	1074.5	15.9	768	4	ABBS9491	Drosophil
14	1074.5	15.9	768	5	AAE25632	Fruit fly
15	1074.5	15.9	788	5	AAU76023	Fruit fly
16	1074.5	15.9	788	5	AAU76014	Fruit fly
17	1074.5	15.9	768	6	ABG72281	Fruit fly
18	443.5	6.5	726	5	AAE25633	Poly aden
19	443.5	6.5	726	5	AAU76024	worm poly
20	443.5	6.5	726	5	AAU76015	worm poly
21	443.5	6.5	726	6	ABG72282	C. elegan
22	350.5	5.2	819	4	ABG20721	Novel hum
23	199	2.9	100	4	ABG11103	Novel hum
24	190	2.8	931	4	ABBG1093	Drosophil
25	186	2.7	33	5	AAE25652	Bovine po
26	178	2.6	2703	4	ABBS60074	Drosophil
27	177.5	2.6	2951	4	ABBS60291	Drosophil
28	174	2.6	718	4	AAAB60140	Human pro
29	172	2.5	1489	6	ABO14779	Novel hum
30	172	2.5	1489	6	ABU11038	Human pro
31	169.5	2.5	1297	6	ABU43768	Protein e
32	169	2.5	699	4	AAW79156	Human pro
33	168	2.5	335	4	ABG09631	Novel hum
34	167	2.5	31	5	AAE25634	Bovine po
35	167	2.5	31	5	AAE25651	Bovine po
36	167	2.5	31	5	AAU76025	Bovine po
37	167	2.5	31	5	AAU76016	Bovine po
38	167	2.5	31	6	ABG72283	Oligopept
39	167	2.5	1169	6	ABU43255	Protein e
40	167	2.5	1177	5	ABP40261	Staphyloc
41	166.5	2.5	688	4	ABG03434	Novel hum
42	166.5	2.5	787	6	ABP55414	Human MDD
43	166	2.5	1187	2	AAE66451	AF-4 prot
44	166	2.5	1210	2	AAE66450	AF-4 prot
45	166	2.5	1210	7	ABR63878	Human AF4

#### ALIGNMENTS

RESULT 1

AAE25631 standard; protein; 968 AA.

XX

AC

AAE25631;

XX

DT 04-NOV-2002 (first entry)

XX

DE Murine poly adenosine diphosphate-ribose glycohydrolase (PARG).

XX

XX

KW Murine; poly adenosine diphosphate-ribose glycohydrolase; PARG; enzyme;

KW ADP-ribose; neoplastic disorder; genotoxic oxidative stress; neurotoxic;

KW cardiac disorder; neuronal disorder; reperfusion injury; neurotoxicity;

KW Alzheimer's disease; Huntington's disease; Parkinson's disease; cardiant;

KW cytosolic; vasotropic; neuroprotective; anticonvulsant; gene therapy;

KW antisense therapy.

XX

OS Mus musculus.

XX

PN US6395543-B1.

XX

PD 28-MAY-2002.

XX

PF 23-FEB-2000; 2000US-00511507.

XX

PR 01-MAY-1998; 98US-0083768P.

PR 30-APR-1999; 99US-00302812.

XX

PA (KENT ) UNIV KENTUCKY RES FOUND.

XX

PI Jacobson MK, Jacobson EL, Ame J, Lin W;

XX

DR WFI; 2002-535641/57.

DR N-PSDE; AAD42083.

XX New nucleic acid molecule encoding bovine poly adenosine diphosphate-  
 PT ribose glycohydrolase involved in cellular response to DNA damage,  
 PT inhibition of which is useful for treating neoplastic disorders and  
 PT neurodegenerative diseases.

XX Claim 3; Col 63-68; 77pp; English.

CC The invention relates to an isolated nucleic acid molecule which encodes  
 CC a bovine poly (adenosine diphosphate (ADP)-ribose) glycohydrolase (DPARG)  
 CC which catalyzes release of ADP-ribose from an ADP ribose polymer. The  
 CC invention is useful as probes and primer molecules that can be used in  
 CC hybridization assays and polymerase chain reaction (PCR) amplification.  
 CC The knowledge of the nucleotide sequence of the PARG gene permits the  
 CC preparation of antisense therapeutics containing sequences complementary  
 CC to the mRNA of PARG gene. The antisense therapeutics are useful to treat  
 CC neoplastic disorders and neurological disorders caused by genotoxic oxidative stress  
 CC e.g., cardiac disorders, neuronal disorders, reperfusion injury,  
 CC neurotoxicity, Alzheimer's disease, Huntington's disease and Parkinson's  
 CC disease. The invention is useful in gene therapy and antisense therapy.  
 CC The present sequence is murine PARG

XX SQ Sequence 968 AA;

Alignment Scores:

Pred. No.: 0 Length: 968  
 Score: 5164.00 Matches: 968  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 76.24% Indels: 0  
 DB: 5 Gaps: 0

US-09-302-812-5 (1-3814) x ABE25631 (1-968)

QY 28 ATGAGTGGCGGCCCCGGCTGGAGCCCTGCAGAAAGCGCTGGCGGCCCGCTGGAAT 87  
 DB 1 MetSerAlaGlyProGlyTrpGluProCysThrLysAlaArgTrpGlyAlaAlaGlyThr 20  
 QY 88 TCTGGCGGCGGCTGGAGTCCCGAGCTCCCTGGCAGGAGGCGGTCTTCGAC 147  
 DB 21 SerAlaProThrAlaSerAspSerArgSerPheProGlyArgGlnArgValLeuAsp 40  
 QY 148 CCCAAGGAGCGCTCCCGCTCAGGTTCAGGTCCCTCGCTCCGCGCCGCTCGGGG 207  
 DB 41 ProLysAspAlaProValGlnPheArgValProProSerSerProAlaCysValSerGly 60  
 QY 208 CGGGCGGAGCCGACAGAGCGAAGCCACCTCGTTGTTTTCACAAAGACTATTA 267  
 DB 61 ArgAlaGlyProHisArgGlyAsnAlaThrSerPheValPheLysGlnLysThrThr 80  
 QY 268 ACTTGGATGGATACTAAAGGACCAAGACAGCTGAATCAGAAAGTAAAGAAACACAT 327  
 DB 81 ThrTrpMetAspThrLysGlyProLysThrAlaGluSerGluSerLysGluAsnAsn 100  
 QY 328 ACAAGAATTGACTCCATGATGAGTCTGTGCAAGAGATACTTTTACCACATAGGTG 387  
 DB 101 ThrArgIleAspSerMetSerValGlnLysAspAsnPheThrProHisLysVal 120  
 QY 388 GAAAAATTGGAAAAATGCTCCTGACGTAATCTGTGATAATCACCACAGAAAGAGTTCA 447  
 DB 121 GluLysLeuGluAsnValProGlnLeuAsnLeuAspLysSerProThrGluLysSer 140  
 QY 448 CAGTATTTCAACCAACAGCAGACTGCGAGTGTGCAAGTGGCAGAAATCAAGGAGCAT 507  
 DB 141 GlnTyLysAsnGlnGlnThrAlaSerValCysLysTrpGlnAsnGluGlyLysHis 160  
 QY 508 GCAGAACAGCTTTTGGCAAGTGAAGCTCCCGGGGAGCTCCGCTACCAAGCAGCTTAGT 567  
 DB 161 AlaGluGlnLeuLeuAlaSerGluProProAlaGlyThrProLeuProLysGlnLeuSer 180  
 QY 568 ATGCTAACATTGGTCACTGACCCACACTGATGACCAAGTGCACACATCATGAAGAA 627  
 DB 181 AsnAlaAsnIleGlyGlnSerProHisThrAspAspHisSerAspThrAspHisGluGlu 200

QY 628 GACAGAGCAATCAGCAGTTCCTTACACCTATAAACTTGCATAACAGCCACAGTA 587  
 DB 201 AspArgAspAsnGlnGlnPheLeuThrProIleLysLeuAlaAsnThrLysProThrVal 220  
 QY 688 GGAGATGGCGAGCCAGCAAGCAACTGTAAGTGCAGTGGATCTCGGCAGTCTGTGAAGAC 747  
 DB 221 GlyAspGlyGlnAlaArgSerAsnCysLysCysSerGlySerArgGlnSerValLysAsp 240  
 QY 748 TGTACAGGCTGTCAACAGAGAGGTGGATGTGCTACAGAGAGTCTTTTGTGATGTT 807  
 DB 241 CysThrGlyCysGlnGlnGluValAspValLeuProGluSerProLeuSerAspVal 260  
 QY 808 GGTGCGGAGGACATGGAAGTGGACCAAAATACAAATTCACATTCAGTGCAGCAAGAAC 867  
 DB 261 GlyAlaGluAspIleGlyThrGlyProLysAsnAspAsnLysLeuThrGlyGlnGluSer 280  
 QY 868 AGCTAGTGTGATTCGCCCTCCATTTGAGAAAGAAAGTGCAGCTGAGTCAACATGATGTA 927  
 DB 281 SerLeuGlyAspSerProPheGluLysGluSerGluProGluSerProMetAspVal 300  
 QY 928 GACAACTCAGAAACAGTTGTCAAGATTCAGAAAGCAGATGAAGAAACAAAGTCCAGTCTTT 987  
 DB 301 AspAsnSerArgAsnSerCysGlnAspSerGluAlaAspGluGluThrSerProValPhe 320  
 QY 988 GATGAGCAAGATGATCGTTCCTCCCAACAGCAAAATAAATTTCAAGTTCACCAAGCA 1047  
 DB 321 AspGluGlnAspAspArgSerSerGlnThrAlaAsnLysLeuSerSerCysGlnAlaArg 340  
 QY 1048 GAACTGTGATGGCGATCTTAGGAAACGGTATTGACTAAGGGAGGTGAAGTTCAGATTGCAT 1107  
 DB 341 GluAlaAspGlyAspLeuArgLysArgTyrLeuThrLysGlySerGluValArgLeuHis 360  
 QY 1108 TTCCAATTGAAGAGAGAAATAATCTGGGACCAAGTCTTAATTAATGCCAAGCCATCTGA 1167  
 DB 361 PheGlnPheGluGlyGluAsnAsnAlaGlyThrSerAspLeuAsnAlaLysProSerGly 380  
 QY 1168 AACTCTTCTAGCTTAATGTTAGAGTGTAGAGTCCAGAGTCCAGCAGCAGTGGAAAGGATCT 1227  
 DB 381 AsnSerSerLeuAsnValGluCysArgSerSerLysGlnHisGlyLysArgAspSer 400  
 QY 1228 AAAATTACAGATCATTTTCATGAGAAATTTCCAAAGTCAGAGGACAGAGAAAGAAACAAT 1287  
 DB 401 LysIleThrAspHisPheMetArgIleSerLysSerGluAspArgArgLysGluGlnCys 420  
 QY 1288 GAACTCAGACATCAAGAACAGAAAGAGATTCCAAATATACATCCACCTAACCTCCCT 1347  
 DB 421 GluValArgHisGlnArgThrGluArgLysIleProLysTyrIleProProAsnLeuPro 440  
 QY 1348 CCAGAGAGAAAGTGGCTGGAACTCTTATTGAGGAAATGAGAAATGCTCGGTGTGG 1407  
 DB 441 ProGluLysLysTrpLeuGlyThrProIleGluGluMetArgLysMetProArgCysGly 460  
 QY 1408 ATCCATTTGCCCTTCCTTAAGACCACTGCAAGTCAACAGTACCTGTCGGGTGACCTT 1467  
 DB 461 IleHisLeuProSerLeuArgProSerAlaSerHisThrValThrValArgValAspLeu 480  
 QY 1468 CTGACAGCAGGAGAGTTCGGAACCTTTTCCACACATTAAGAGATTTGTGGGATAC 1527  
 DB 481 LeuArgAlaGlyGluValProLysProPheProThrHisTyrLysAspLeuTrpAspAsn 500  
 QY 1528 AAACATGTGAAATTCCTTGTTCGGAACAAAATCTGTACCTCTGTGGAAGATGAGAAATG 1587  
 DB 501 LysHisValLysMetProCysSerGluGlnAsnLeuTyrProValGluAspGluAsnGly 520  
 QY 1588 GAGCAACTGACGAGGAGTGTGGAGCTCATTCAGACTGCTTCCTCAACAAATTCACA 1647  
 DB 521 GluArgThrAlaGlySerArgTrpGluLeuIleGlnThrAlaLeuLeuAsnLysPheThr 540  
 QY 1648 CGACCCCAAGACTTGAAGGATGCGATTCCTGAAATACATGTCGATATCTTAAGAAATGG 1707  
 DB 541 ArgProGlnAsnLeuLysAspAlaIleLeuLysTyrAsnValAlaTyrSerLysLysTrp 560

QY 1708 GACCTTACAGCTTGGTGTGATTTCTGGATAAGCTACTTGAAGAAGCAGAGGCCCAACAT 1767  
Db 561 AspPheThrAlaLeuValAspPheTrpAspLysValLeuGluGluAlaGluAlaHis 580  
QY 1768 TTATATCAGTCCATTTTACCTGCATGGTGAATAATTCACCTCTGTCTGCGAAATATTTGC 1827  
Db 581 LeuTyrGlnSerIleLeuProAspMetValLysIleAlaLeuCysLeuProAsnIleCys 600  
QY 1828 ACCCAGCCATACCTCTGAAACAGAGATCAATCATTCTCTCAGATGTCACAGGAA 1887  
Db 601 ThrGlnProLeuLeuLysGlnLysMetAsnHisSerValThrMetSerGlnGlu 620  
QY 1888 CAGATGCCAGCTTTTGTAGCTAATAGCTTTCTTCTGCACATTTTCCCGCAGCGAATGCCAAG 1947  
Db 621 GlnIleAlaSerLeuLeuAlaAsnAlaPhePheCysThrPheProArgArgAsnAlaLys 640  
QY 1948 ATGAAATCGAGTATTTCTAGTACCCAGACATTAATCTCAATCGGTGTTTGAAGACGT 2007  
Db 641 MetLysSerGluTyrSerSerTyrProAspIleAsnPheAsnArgLeuPheGluGlyArg 660  
QY 2008 TCATCAAGGAACCCAGAAACCTGAAACACTCTTCTGTACTTTCGAAAGTCAAGAG 2067  
Db 661 SerSerArgLysProGlnLysLeuLysThrLeuPheCysTyrPheArgargValThrGlu 680  
QY 2068 AAAAAACCTACAGGATTGGTGACATTTTACAAGACAGAGTCTTGAAGATTTTCCAGAAATGG 2127  
Db 681 LysLysProThrGlyLeuValThrPheThrArgGlnSerLeuGluAspPheProGluTrp 700  
QY 2128 GAAAGTGTGAAAGCTCTGACACGCTTACAGCTCATTACAGGGTACCATAGAGGC 2187  
Db 701 GluArgCysGluLysProLeuThrArgLeuHisValThrTyrGluGlyThrIleGluGly 720  
QY 2188 AACGGCCGAGGATGCTACAGGTGGATTTTGGCAATCGTTTGTGAGTGTGTGACT 2247  
Db 721 AsnGlyArgGlyMetLeuGlnValAspPheAlaAsnArgPheValGlyGlyValThr 740  
QY 2248 GGTGCGGAGCTTGTACAAGAAGAAATCAGATTTTAAATCAATCCTGAAATGATGTTTCA 2307  
Db 741 GlyAlaGlyLeuValGlnGluGluIleArgPheLeuIleAsnProGluLeuIleValSer 760  
QY 2308 CGGCTGTTCTCAGGTGCTGGATCAATGAGTGTCTTATATATCAGTACTGAAACAG 2367  
Db 761 ArgLeuPheThrGluValLeuAspHisAsnGluCysLeuIleIleThrGlyThrGluGln 780  
QY 2368 TACAGTGAATACACAGCTATGCTGAAACTTATCTGTTGGCCGAGCCATGAAGATGG 2427  
Db 781 TyrSerGluTyrThrGlyTyrAlaGluThrTyrArgTrpAlaArgSerHisGluAspGly 800  
QY 2428 AGTGAAGAAGCAGATTGGCAGCGCGCTGCACGGAGATCGTTGCCATTGACGCACTTCAC 2487  
Db 801 SerGluLysAspAspTrpGlnArgArgCysThrGluIleValAlaIleAspAlaLeuHis 820  
QY 2488 TTCAGACGCTACTCCTCATCAGTTTGTGCTGAGAAAGTGAGAGTCAAGCTTAACAGGCT 2547  
Db 821 PheArgArgTyrLeuAspGlnPheValProGluLysValArgargGluLeuAsnLysAla 840  
QY 2548 TACTCGGATTCCTCGTCTCGAGTTCCTTCTGAAAATCTTTCTCAGTGGCCACGGGA 2607  
Db 841 TyrCysGlyPheLeuArgProGlyValProSerGluAsnLeuSerAlaValAlaThrGly 860  
QY 2608 AACTGGGCTGTGGTGCCTTTGGGGGTGACGCTAGATTAAAGCCTTAATPACAGATCCTG 2667  
Db 861 AsnTrpGlyCysGlyAlaPheGlyGlyAspAlaArgLeuLysAlaLeuIleGlnIleLeu 880  
QY 2668 GCAGCTGCTCGGCTGAACGTGACGTGGTATTATTCCTTTGGGACTCAGATTTGATG 2727  
Db 881 AlaAlaAlaAlaAlaGluArgAspValValTyrPheThrPheGlyAspSerGluLeuMet 900  
QY 2728 AGAGACATTTACAGCATGCACATTTCTTACCAGAGGAGGAGTGTGATGTTGAAAAGTG 2787  
Db 901 ArgAspIleTyrSerMetHisThrPheLeuThrGluArgLysLeuAspValGlyLysVal 920  
QY 2788 TACAAGTTATTGCTTAGATACTAATGAAGAATGCAGAAACTGTTTCCACCCCTGGACCA 2847

Db 921 TyrLysLeuLeuLeuArgTyrTyrAsnGluGluCysArgAsnCysSerThrProGlyPro 940  
QY 2848 GACATCAAGCTTTATCCATTATACCATGCTGTTGAGTCAAGTGCAGAGACCACTGAC 2907  
Db 941 AspIleLysLeuTyrProPheIleTyrHisAlaValGluSerSerAlaGluThrThrAsp 960  
QY 2908 ATGCCAGGACAGAAGGCAGGCACC 2931  
Db 961 MetProGlyGlnLysAlaGlyThr 968  
RESULT 2  
AAU76022  
ID AAU76022 standard; protein; 968 AA.  
XX AAU76022;  
XX  
XX 08-MAY-2002 (first entry)  
XX Mouse poly(ADP-ribose) glycohydrolase (PARG) protein sequence.  
XX Mouse; poly(ADP-ribose) glycohydrolase; PARG; DNA repair; ADP-ribose;  
KW adenosine diphosphate-ribose; neoplastic disorder; adenocarcinoma;  
KW leukaemia; lymphoma; melanoma; myeloma; sarcoma; teratocarcinoma;  
KW hyperplasia; hypertrophy; reperfusion; ischaemia; heart attack; stroke;  
KW neurodegenerative disease; neurological disorder; Alzheimer's disease;  
KW Huntington's disease; Parkinson's disease.  
XX  
XX Mus musculus.  
XX  
XX US6337202-B1.  
XX  
XX 08-JAN-2002.  
XX  
XX 23-FEB-2000; 2000US-00511477.  
XX  
XX 01-MAY-1998; 98US-0083768P.  
XX 30-APR-1999; 99US-00302812.  
XX  
XX (KENT ) UNIV KENTUCKY RES FOUND.  
XX Jacobson MK, Jacobson EL, Ame J, Lin W;  
XX WPI; 2002-163240/21.  
XX N-PSDB; ABK14933.  
XX  
XX Novel isolated and purified poly (ADP-ribose) glycohydrolase protein  
PT which catalyses release of ADP-ribose from ADP-ribose polymer, useful for  
PT treating neoplastic and neurological disorders, heart attack and stroke.  
XX  
XX Claim 2; Col 63-70; 81pp; English.  
XX  
XX The present invention relates to a new poly(ADP-ribose) glycohydrolase  
CC (PARG) protein which catalyses release of ADP-ribose from an ADP  
CC (adenosine diphosphate)-ribose polymer. The PARG molecule of the  
CC invention is useful for generating antibodies and can be inhibited or  
CC activated for diagnosing and treating neoplastic disorders such as  
CC adenocarcinoma, leukaemia, lymphoma, melanoma, myeloma, sarcoma,  
CC teratocarcinoma, hyperplasia and hypertrophy, reperfusion following  
CC ischaemia, heart attack, stroke, neurodegenerative diseases, neurological  
CC disorders including Alzheimer's, Huntington's and Parkinson's diseases,  
CC and related conditions. PARG levels may be enhanced to suppress DNA  
CC repair and increase the cell's susceptibility to chemotherapy drugs.  
CC Antagonists of PARG are administered to treat or prevent neoplastic  
CC disorders. The present amino acid sequence represents the mouse PARG  
CC protein of the invention. This protein is one of several PARG proteins  
XX (AAU76020-AAU76024) of the invention  
XX  
XX Sequence 968 AA;  
Alignment Scores: 0 Length: 968  
Pred. No.: 5164.00 Matches:  
Score:



Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	76.24%	Indels:	0
DB:	5	Gaps:	0
US-09-302-812-5 (1-3814) x AAU76022 (1-968)			
QY	28	ATGAGTGGCGGCCCCGGCTGGAGCCCTGCACGAAGAGCGCGCTGGGGCCCGCTGGAAT	87
DB	1	MetSerAlaGlyProGlyTrpGluProCysThrLysAlaArgTrpGlyAlaAlaGlyThr	20
QY	88	TCTGGCGCCAGCTGCCCTCGGACTCCCGAGCTCCCTCGGAGGAGGAGGCTGTCTCGAC	147
DB	21	SerAlaProThrAlaSerAspSerArgSerPheProGlyArgGlnArgValLeuAsp	40
QY	148	CCCAAGAGCGCTCCCGTCCAGTTCCAGTTCAGGTCCTCCGCTCCGCGAGCTGCTCGGGG	207
DB	41	ProLysAspAlaProValGlnPheArgValProProSerSerProAlaCysValSerGly	60
QY	208	CGGGCGGACCCGACAGAGGCAACGCCACCTCGTTGTTTTCAAAACAAAGCTATTACT	267
DB	61	ArgAlaGlyProHisArgGlyAsnAlaThrSerPheValPheLysGlnLysThrIleThr	80
QY	268	ACTTGATCGATCTAAAGGACCAAGACAGCTGAATCAGAAAGTAAAGAAACAACAAT	327
DB	81	ThrTrpMetAspThrLysGlyProLysThrAlaGluSerGluSerLysGluAsnAsn	100
QY	328	ACAAGAATTGATCCCATGATGATGTTCTGTGAGAAAGATACTTTTACCCACATAAGGTG	387
DB	101	ThrArgIleAspSerMetMetSerValGlnLysAspAsnPheTyProHisLysVal	120
QY	388	GAAAATTTGAAAATGTTCTCAGCTAACTCTGATTAATCAACCCACAGAAAGAGTTCA	447
DB	121	GluLysLeuGluAsnValProGlnLeuAsnLeuAspLysSerProThrGluLysSer	140
QY	448	CAGTATTTGAACCAACAGCAGACTCGAGTGTGCAAGTGCAGATGAAGGAGCAT	507
DB	141	GlnTyLeuAsnGlnGlnThrAlaSerValCysLysTrpGlnAsnGluGlyLysHis	160
QY	508	GCAGAACAGCTTTGGCAAGTGAAGCTCCCGGGGAGCTCCGCTACCAAGCAGCTTAGT	567
DB	161	AlaGluGlnLeuLeuAlaSerGluProProAlaGlyThrProLeuProLysGlnLeuSer	180
QY	568	ATGCTTAACATTTGTCAGTCCACCCACACTGATGACCAAGTGCAGATGACATCATGAG	627
DB	181	AsnAlaAsnIleGlyGlnSerProHisThrAspAspHisSerAspThrAspHisGluGlu	200
QY	628	GACAGAGACAATTCAGCAGTTTCTTACACCTATAAACTTGCAAAATCAAGCAACAGTA	687
DB	201	AspArgAspAsnGlnGlnPheLeuThrProLysLeuAlaAsnThrLysProThrVal	220
QY	688	GGAGATGGGCGGCGAGAGCAACTGTAGTGCATGATCTCGCAGTCTGTGAAAGAC	747
DB	221	GlyAspGlyGlnAlaArgSerAsnCysLysSerGlySerArgGlnSerValLysAsp	240
QY	748	TGTACAGGCTGTCACAGAGAGAGTGGTGTGCTACAGAGAGTCCCTTGTTCAGATGT	807
DB	241	CysThrGlyCysGlnGlnGluValAspValLeuProGluSerProLeuSerAspVal	260
QY	808	GGTGGCGGAGCATTTGGAACTCGACAAATAATGACAAATTCAGCTGGACAAAGAGC	867
DB	261	GlyAlaGluAspIleGlyThrGlyProLysAsnAspAsnLysLeuThrGlyGlnGluSer	280
QY	868	AGCTAGTGTATTCGCTCCATTTGAGAAAGAAAGTGAAGCTGAGTCCCAATGATGTA	927
DB	281	SerLeuGlyAspSerProProPheGluLysGluSerGluProGluSerProMetAspVal	300
QY	928	GACAACTCGAGAAACAGTTGTCAAGATTCAGAAAGCAGATGAGAAACAGTCCAGTCTTT	987
DB	301	AspAsnSerArgAsnSerCysGlnAspSerGluAlaAspGluGluThrSerProValPhe	320
QY	988	GATGAGCAAGATGATCGTTCTCCCAACAGCAAAATAAACTTCAAGTTGCCCAAGACA	1047
DB			
DB	321	AspGluGlnAspAspArgSerSerGlnThrAlaAsnLysLeuSerSerCysGlnAlaArg	340
QY	1048	GAAAGTGTATGGCGATCTTTAGGAAACGGTATTTGACTTAAGGGAAGTGAATTTAGAT	1107
DB	341	GluAlaAspGlyAspLeuArgLysArgTyLeuThrLysGlySerGluValArgLeuHis	360
QY	1108	TTCCAAATTTGAAGGAGAAATTAATCTCTGGACCACTGACTTAATGCCAGCACTGGA	1167
DB	361	PheGlnPheGluGlyGluAsnAsnAlaGlyThrSerAspLeuAsnAlaLysProSerGly	380
QY	1168	AACTCTTCTAGCCTTAAATGTAGAGTGTAGAACTTCCAAGCAGCATGGAAAAAGGATCT	1227
DB	381	AsnSerSerSerLeuAsnValGluCysArgSerSerLysGlnHisGlyLysArgAspSer	400
QY	1228	AAAATAAGATCATTTTCATGAGAAATTTCCAAAGTCAGAGGACAGAAAGAAACAATGT	1287
DB	401	LysIleThrAspHisPheMetArgLysSerLysSerGluAspArgLysGluGlnCys	420
QY	1288	GAAAGTCACACATCAAGAACAGAAAGAGAGATTCCAAATATCATCCACCTAACCTCCCT	1347
DB	421	GluValArgHisGlnArgThrGluArgLysIleProLysTyIleProProAsnLeuPro	440
QY	1348	CCAGAGAAAGAGTGGCTGGGAACCTCTATTGAGGAAATGAGAAAAATCGCTCGGTGGG	1407
DB	441	ProGluLysLysTrpLeuGlyThrProIleGluGluMetArgLysMetProArgCysGly	460
QY	1408	ATCCATTTGCCCTTCTTAAGACCATCTGCAAGTGCACAGAGTCTGTCGGGTGACCTT	1467
DB	461	IleHisLeuProSerLeuArgProSerAlaSerHisThrValThrValArgValAspLeu	480
QY	1468	CTGAGAGCAGAGAGGTTCGAAACCTTTTCCAAACACATTAACAAGATTTGGGATAC	1527
DB	481	LeuArgAlaGlyGluValProLysProPheProThrHisTyLysAspLeuTrpAspAsn	500
QY	1528	AAACATGTGAAATTCCTTGTTCGGAACAAAATCTGTACCTCTGTGGAAGATGAGAAATGT	1587
DB	501	LysHisValLysMetProCysSerGluGlnAsnLeuTyProValGluAspGluAsnGly	520
QY	1588	GAGCGAACTGCAGGAGTAGTGGAGCTCAATCAGACTGCATCTCTCAACAAATTCACA	1647
DB	521	GluArgThrAlaGlySerArgTrpGluLeuIleGlnThrAlaLeuLeuAsnLysPheThr	540
QY	1648	CGACCCAGAACTTCAGGATGCGATCTGAAATACATATGAGTGGCATATTTCTAAGAAATGG	1707
DB	541	ArgProGlnAsnLeuLysAspAlaIleLeuLysTyAsnValAlaTySerLysLysTrp	560
QY	1708	GACTTTACAGCTTTGGTTGATTTCTGGGATAAGGTACTTGAAGAGCAGAGCCCAACAT	1767
DB	561	AspPheThrAlaLeuValAspPheTrpAspLysValLeuGluGluAlaGlnHis	580
QY	1768	TTATATCAGTCCATTTTACCTGCACATGGTGAATTAATTCATCTGTCTGCCAAATTTTGC	1827
DB	581	LeuTyGlnSerIleLeuProAspMetValLysIleAlaLeuCysLeuProAsnIleCys	600
QY	1828	ACCAGCCAAATACCACTCTCAACAGAGAGATGAATCATTTCTGCACCATGTCACAGAA	1887
DB	601	ThrGlnProIleProLeuLeuLysGlnLysMetAsnHisSerValThrMetSerGlnGlu	620
QY	1888	CAGATCGCCAGCTCTTTAGCTTAATGCTTTCTCTCTGTCACATTTCCCGCAGGAATGCCAAG	1947
DB	621	GlnIleAlaSerLeuLeuAlaAsnAlaPhePheCysThrPheProArgArgAsnAlaLys	640
QY	1948	ATGAAATCGAGATTTTATGTTACCCAGACATTAATCAATCGTGTGTTTGAAGGACGT	2007
DB	641	MetLysSerGluTySerSerTyProAspIleAsnPheAsnArgLeuPheGluGlyArg	660
QY	2008	TCATCAAGGAACACAGAAAACTGAAACACTCTTCTGCTACTTTTCGAAGAGTCACAGAG	2067
DB	661	SerSerArgLysProGluLysLeuLysThrLeuPheCysTyThrPheArgArgValThrGlu	680
QY	2068	AAAAAACTACAGGATTTGGTGCATTTTACAAGACAGAGTCTTGAAGATTTTCCAGATGG	2127
DB	681	LysLysProThrGlyLeuValThrPheThrArgGlnSerLeuGluAspPheProGluTrp	700

QY 2128 GAAAGGTGTGAAAGCCTCTGACACGCTTACACCTCTACAGGCTTACATAGAGGC 2187  
 Db |||||  
 QY 701 GluArgCysGluLysProLeuThrArgLeuHisValThrGluGlyThrIleGluGly 720  
 Db |||||  
 QY 2188 AACGCCGAGCATGCTACAGGTGGATTTGCAATCGTTTGTGGAGGTGGTGAAT 2247  
 Db |||||  
 QY 721 AsnGlyArgGlyMetLeuGlnValAspPheAlaAsnArgPheValGlyGlyValThr 740  
 Db |||||  
 QY 2248 GTCGGCGGACTGTACAGAAGAATCAGATTTTATCAATCTGTAATGATGTTTCA 2307  
 Db |||||  
 QY 741 GlyAlaGlyLeuValGlnGluGluIleArgPheLeuIleAsnProGluLeuIleValSer 760  
 Db |||||  
 QY 2308 CGGCTGTTCACTGAGGTGCTGATCAATGAGTGTCTTATATCAAGCTACTGACAG 2367  
 Db |||||  
 QY 761 ArgLeuPheThrGluValLeuAspHisAsnGluCysLeuIleIleThrGlyThrGluGln 780  
 Db |||||  
 QY 2368 TACAGTGAATACACAGGCTATGCTGAAACTTATCGTTGGGCCCGAAGCCATGAAGATGGG 2427  
 Db |||||  
 QY 781 TyrSerGluThrThrGlyTyrAlaGluThrTyrArgTrpAlaArgSerHisGluAspGly 800  
 Db |||||  
 QY 2428 AGTGAAAGACAGATTGCGAGCGCGCTGACGAGATCGTGTCCATGACGCACTTAC 2487  
 Db |||||  
 QY 801 SerGluLysAspAspTrpGlnArgArgCysThrGluIleValAlaIleAspAlaLeuHis 820  
 Db |||||  
 QY 2488 TTCAGACGCTACCTCGATCAGTTTGTCCCTGAGAACTGAGCTGACGTGACCTTAAAGGCT 2547  
 Db |||||  
 QY 821 PheArgArgTyrLeuAspGlnPheValProGluLysValArgGluLeuAsnLysAla 840  
 Db |||||  
 QY 2548 TACTGCGGATTCCTCCCTCGAGTTCCTTCTGAAATCTTCTGAGTGGCCACCGGGA 2607  
 Db |||||  
 QY 841 TyrCysGlyPheLeuArgProGlyValProSerGluAsnLeuSerAlaValAlaThrGly 860  
 Db |||||  
 QY 2608 AACTGGGCTGTGCTGCTTGGGGTGACGTAGATTAAGCTTAAATACAGATCCTG 2667  
 Db |||||  
 QY 861 AsnTrpGlyCysGlyAlaPheGlyGlyAspAlaArgLeuLysAlaLeuIleGlnIleLeu 880  
 Db |||||  
 QY 2668 GCAGCTGCTGCGCTGAACGTGAGTGTGTTATTTACCTTGGGACTCAGAGTTGATG 2727  
 Db |||||  
 QY 881 AlaAlaAlaAlaAlaGluArgAspValValTyrPheThrPheGlyAspSerGluLeuMet 900  
 Db |||||  
 QY 2728 AGAGACATTTACAGCATGCACATTTCTTCCCTTACCGAGAGGAGCTGGATGTGTGAAAGTG 2787  
 Db |||||  
 QY 901 ArgAspIleTyrSerMetHisThrPheLeuThrGluArgLysLeuAspValGlyLysVal 920  
 Db |||||  
 QY 2788 TACAAGTTATGCTAGATACATCAATGAAGAATGCAAACTGTTCACCCCTGGACCA 2847  
 Db |||||  
 QY 921 TyrLysLeuLeuLeuArgTyrTyrAsnGluGluCysArgAsnCysSerThrProGlyPro 940  
 Db |||||  
 QY 2848 GACATCAAGCTTTATCCATTATATCATATCATGCTGTGTTGAGTCAAGTGCAGAGACCTGAC 2907  
 Db |||||  
 QY 941 AspIleLysLeuTyrProPheIleTyrHisAlaValGluSerSerAlaGluThrThrAsp 960  
 Db |||||

RESULT 3

AAU76013

ID AAU76013 standard; protein; 968 AA.

AC

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

OS Mus musculus.  
 XX US6333148-B1.  
 XX 25-DEC-2001.  
 PD 30-APR-1999; 99US-00302812.  
 XX 01-MAY-1998; 98US-0083768P.  
 XX (KENT ) UNIV KENTUCKY RES FOUND.  
 XX Jacobson MK, Jacobson EL, Ame J, Lin W;  
 DR WPI; 2002-153820/20.  
 DR N-PSDB; ABK14495.  
 XX  
 PT Screening compounds for modulation of poly(ADP-ribose) glycohydrolase,  
 PT useful potentially for treating diseases associated with DNA damage, e.g.  
 PT cancer.  
 XX  
 PS Claim 3; Col 63-68; 80pp; English.  
 XX  
 CC The present invention relates to a new method for screening compounds for  
 CC ability to modulate activity of an enzyme that hydrolyses ADP (adenosine  
 CC diphosphate)-ribose from an ADP-ribose polymer. The compounds are  
 CC inhibitors or activators of PARG (poly(ADP-ribose) glycohydrolase) and  
 CC are used to treat or prevent any condition associated with DNA damage  
 CC e.g. neoplasia, inherited genetic diseases, myocardial infarction,  
 CC vascular stroke, aging and neurodegeneration e.g. Huntington's,  
 CC Parkinson's or Alzheimer's diseases, or neurotoxicity generally.  
 CC Compounds identified by the new method are more effective than known  
 CC inhibitors and have fewer side effects. The present amino acid sequence  
 CC represents the mouse PARG protein of the invention. This protein is one  
 CC of several PARG proteins (AAU75799 and AAU76012-AAU76015) of the  
 CC invention  
 XX

SQ Sequence 968 AA;

Alignment Scores:

Pred. No.: 0 Length: 968  
 Score: 5164.00 Matches: 968  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 76.24% Indels: 0  
 DB: 5 Gaps: 0

US-09-302-812-5 (1-3814) x AAU76013 (1-968)

QY 28 ATGAGTGGCGGCCCCCGCTGGGAGCCCTGCACGAAGCGGCTGGGCGCGCTGGNACT 87  
 Db |||||  
 1 MetSerAlaGlyProGlyTrpGluProCysThrLysAlaAargTrpGlyAlaAlaGlyThr 20  
 QY 88 TCTGCGCGGACTGCTCGGACTCCCGAGCTTCCCTGGCGAGGAGGCGGTGTTCTCGAC 147  
 Db |||||  
 21 SerAlaProThrAlaSerAspSerArgSerPheProGlyArgGlnArgValLeuAsp 40  
 QY 148 CCCAAGGACGCTCCCGCTCCAGTTCAGGTCCCTCCGCTCCGCGAGCTGCGTCTCGGG 207  
 Db |||||  
 41 ProLysAspAlaProValGlnPheArgValProProSerSerProAlaCysValSerGly 60  
 QY 208 CGGGCGGAGCCGACAGAGGCAACCCACCTGTTGTTTTCACAAAGAGACTATTACT 267  
 Db |||||  
 61 ArgAlaGlyProHisArgGlyAsnAlaThrSerPheValPheLysGlnLysThrIleThr 80  
 QY 268 ACTTGGATGGATCTAAAGGACCCCAAGACAGCTGATCAGAAAGTAAAGAAACCAAT 327  
 Db |||||  
 81 ThrTrpMetAspThrLysGlyProLysThrAlaGluSerGluSerLysGluAsnAsn 100  
 QY 328 ACAAGATTGATCCCATGATGATGTTCTGTGCGAAGAGATACTTTTACCACATAGGTG 387  
 Db |||||  
 101 ThrArgIleAspSerMetMetSerValGlnLysAspAsnPheTyrProHisLysVal 120

Mouse poly(ADP-ribose) glycohydrolase (PARG) protein sequence.

Mouse; poly(ADP-ribose) glycohydrolase; PARG; PARG modulator; ADP-ribose;  
 adenosine diphosphate-ribose; DNA repair; apoptosis; neoplasia; cancer;  
 inherited genetic disease; myocardial infarction; vascular stroke; aging;  
 neurodegeneration; Huntington's disease; Parkinson's disease;  
 Alzheimer's disease; neurotoxicity.

QY 388 GAAAAATTGGAAATGTTCTTCAGCTAAATCTTGATAAATCAACCCACAGAAAGAGTTCA 447  
DB 121 GluLysLeuGluAasnValProGlnLeuAasnLeuAaspLysSerProThrGluLysSerSer 140  
QY 448 CAGTATTTGAAACACAGACGACTCGGAGTGTCGACAGTGGCAGAAAGAGGAGCAT 507  
DB 141 GlnTyrLeuAasnGlnGlnThrAlaSerValCysLysTyrPglAasnGluGlyLysHis 160  
QY 508 GCAGAACAGCTTTTGGCAAGTGAGCTCCCGGGGAGCTCCGCTACCAAGCAGCTTAGT 567  
DB 161 AlaGluGlnLeuLeuAalaserGluProProAlaGlyThrProLeuProLysGlnLeuSer 180  
QY 568 AATGCTAATCTGTCAGTCAACCCACACAGTGAACACAGTGAACAGATCATGAAGAA 627  
DB 181 AasnAlaasnIleGlyGlnSerProHisThrAaspAaspHisSerAaspThrAaspHisGluGlu 200  
QY 628 GACAGAGCAATCAGCAGTCTTACACCTATAAACAATTTGCAAAATACAAAGCCACACTA 687  
DB 201 AspArgAaspAasnGlnGlnPheLeuThrProIleLysLeuAlaasnThrLysProThrVal 220  
QY 688 GGAGATGGCGCGGCGCAGAGCAACTGTAAAGTCAGTGGATCTCGCCAGTCTGTGAAGAC 747  
DB 221 GlyAaspGlyGlnAlaAargSerAasnCysLysCysSerGlySerArgGlnSerValLysAasp 240  
QY 748 TGTACAGGCTGTCAACAGAGAGGTGGATGTGTACAGAGAGTCTTTGTTCAGATGTT 807  
DB 241 CysThrGlyCysGlnGlnGluValAaspValLeuProGluSerProLeuSerAaspVal 260  
QY 808 GGTGCGCAGGACATTGGAACTGGACCAAAAATGACAAATGACATGACGACAGCAAGC 867  
DB 261 GlyAlaGluAaspIleGlyThrGlyProLysAasnAaspAasnLysLeuThrGlyGlnLysSer 280  
QY 868 AGCCTAGTGTATTCGCTCCATTTGAGAAAGAGTGAAGTGAAGTCAACCAATGATGATA 927  
DB 281 SerLeuGlyAaspSerProPheGluLysGluSerGluProGluSerProMetAaspVal 300  
QY 928 GACAACTCGAGAAACAGTTGTCAAGATTCAGAGCAGATGAAGAAACAAAGTCCAGTCTTT 987  
DB 301 AaspAasnSerArgAasnSerCysGlnAaspSerGluAlaAaspGluThrSerProValPhe 320  
QY 988 GATCAGCAAGATGATCGTTCCTCCCAACACACAAATAAATTTCAAGTTGCCAAGCAAGA 1047  
DB 321 AspGluGlnAaspAaspArgSerSerGlnThrAlaasnLysLeuSerCysGlnAlaAarg 340  
QY 1048 GAAGCTGATGCGGATCTTAGAAACGGTATTGTGACTAAGGAAAGTGAAGTTAGATTGAT 1107  
DB 341 GluAlaAaspGlyAaspLeuAargLysArgTyrLeuThrLysGlySerGluValAargLeuHis 360  
QY 1108 TTCAATTTGAAGAGAAATAATGCTGGGACCGAGTGAATTAATGCCAAGCCATCTGGA 1167  
DB 361 PheGlnPheGluGlyGluAasnAlaGlyThrSerAaspLeuAasnAlaLysProSerGly 380  
QY 1168 AACTCTTCTAGCCTTAATGTAGAGTGTAGAAAGTTCCAGCAGCAGATGGAAGAAAGGATTCT 1227  
DB 381 AasnSerSerSerLeuAasnValGluCysAargSerSerLysGlnHisGlyLysAargAaspSer 400  
QY 1228 AAAATTACAGATCATTTTCATGAGAAATTTCCAAAGTCAGAGGACAGAGAAAGAAACAATGT 1287  
DB 401 LysIleThrAaspHisPheMetArgIleSerLysSerGluAaspArgLysGluGlnCys 420  
QY 1288 GAAGTCAGACATCAAGACAGAAAGAGNAGATTCCAAATATCATCCCACTTAACCTCCCT 1347  
DB 421 GluValArgHisGlnAargThrGluAargLysIleProLysTyrIleProProAasnLeuPro 440  
QY 1348 CCAGAGAAAGAGTGGCTGGGAACCTTATTGAGGAAATGAGAAATGCTCGGTGGG 1407  
DB 441 ProGluLysLysTyrLeuGlyThrProIleGluGluMetAargLysMetProArgCysGly 460  
QY 1408 ATCAATTTGCTTCTTAAAGACCATCTGCAAGTCAACAGTCACTGTTCGGGTAGACCTT 1467  
DB 461 IleHisLeuProSerLeuAargProSerAlaSerHisThrValThrValAargValAaspLeu 480  
QY 1468 CTGAGACGAGAGAGTTCCGAAACCTTTTCCACACATTAACAAGATTGTGGGATAAC 1527

DB 481 LeuArgAlaGlyGluValProLysProPheProThrHisTyrLysAaspLeuTyrAaspAasn 500  
QY 1528 AAAATGTGAAATGCTTGTTCGGAACAAACTGTATCCCTGGAGAGATGAGATGGT 1587  
DB 501 LysHisValLysMetProCysSerGluGlnAasnLeuTyrProValGluAaspGluAasnGly 520  
QY 1588 GAGCGAACTGCAGGAGTAGTGGAGCTCAATTCAGACTGCACCTTCTCAACAAATTCACA 1647  
DB 521 GluArgThrAlaGlySerArgTyrGluLeuIleGlnThrAlaLeuLeuAasnLysPheThr 540  
QY 1648 CGACCCAGAACTTGAAGATGCGATCTGAAATACAAATGTGGCATATTCTAAGAAATGG 1707  
DB 541 ArgProGlnAasnLeuLysAaspAlaIleLeuLysTyrAasnValAlaTyrSerLysLysTyr 560  
QY 1708 GACTTTACAGCTTTGGTTGATTTCTGGGATAAGTACTTGAAGAGCAGAGCCCAACAT 1767  
DB 561 AspPheThrAlaLeuValAaspPheTyrAaspLysValLeuGluGluAlaGluAlaGlnHis 580  
QY 1768 TTATATCAGTCCATTTTACCTGACATGTTGAAAAATTGCACCTCTGTCTGCCAAATATTTC 1827  
DB 581 LeuTyrGlnSerIleLeuProAaspMetValLysIleAlaLeuCysLeuProAasnIleCys 600  
QY 1828 ACCCAGCAATACCACTCTGAAACAGAGATGAATCATTTCTGTCAAGATGTCAAGAA 1887  
DB 601 ThrGlnProIleProLeuLeuLysGlnLysMetAasnHisSerValThrMetSerGlnGlu 620  
QY 1888 CAGATCCCGAGTCTTTAGTCTTCTTCGACATTTCCCGACATTTCCCGACGAATGCCAAG 1947  
DB 621 GlnIleAlaSerLeuLeuAlaAasnAlaPhePheCysThrPheProAargAargAasnAlaLys 640  
QY 1948 ATGAAATCGAGATATTCTAGTTACCCAGACATAACTTCAATCGTGTGTTTGAAGAGCGT 2007  
DB 641 MetLysSerGluTyrSerSerTyrProAaspIleAasnPheAasnArgLeuPheGluGlyArg 660  
QY 2008 TCATCAGGAAACACAGAAACACTGAAACACTCTTCTGTACTTTCGAGAGTCAACAG 2067  
DB 661 SerSerArgLysProGluLysLeuLysThrLeuPheCysTyrPheArgArgValThrGlu 680  
QY 2068 AAAAACTCAGGATGTTGACATTACAGACAGAGTCTTGAAGATTTTCCAGAAATGG 2127  
DB 681 LysLysProThrGlyLeuValThrPheThrArgGlnSerLeuGluAaspPheProGluTyr 700  
QY 2128 GAAAGTGTGAAAGCTCTGACACGCTTACAGTCACTTACAGAGGTACATAGAGGC 2187  
DB 701 GluArgCysGluLysProLeuThrArgLysHisValThrTyrGluGlyThrIleGluGly 720  
QY 2188 AACGCCGAGGCATGCTACAGTGGATTTTGCAAATCGTGTGTTTGGAGGTGGTGAAT 2247  
DB 721 AasnGlyArgGlyMetLeuGlnValAaspPheAlaAasnArgPheValGlyGlyValThr 740  
QY 2248 GGTGCGGAGCTTGTACAAAGAAAGAAATCAGATTTTAAATCAATCTCAATGATTGTTTCA 2307  
DB 741 GlyAlaGlyLeuValGlnGluGluIleAargPheLeuIleAasnProGluLeuIleValSer 760  
QY 2308 CGGCTGTTCTGAGGTGCTGGATCAATGAGTGTCTTATATATCAAGGTACTGAGACAG 2367  
DB 761 ArgLeuPheThrGluValLeuAaspHisAasnGluCysLeuIleThrGlyThrGluGln 780  
QY 2368 TACAGTGAATACACAGCTATGCTGAACCTATCGTTGGCCGAGCCATGAGATGGG 2427  
DB 781 TyrSerGlyTyrThrGlyTyrAlaGluThrTyrArgTyrAlaArgSerHisGluAaspGly 800  
QY 2428 AGTGAAGACACATTTGGCAGCGCTGCACGAGATCGTTGCCATTGACGACCTTCAC 2487  
DB 801 SerGluLysAaspTyrGlnAargAargCysThrGluIleValAlaIleAaspAlaLeuHis 820  
QY 2488 TTCAGACGCTTACCTCGATCAGTTGTGCTCGAGAAAGTGAAGTGAAGTGAAGCTTACAGGCT 2547  
DB 821 PheArgArgTyrLeuAaspGlnPheValProGluLysValArgArgGluLeuAasnLysAla 840  
QY 2548 TACTCGGATTCCTCCGTCTCGAGTCTCTCTGAAATCTTCTTCAGTGGCCACCGGA 2607

841 TyrCysGlyPheLeuArgProGlyValProSerGluAsnLeuSerAlaValAlaThrGly 860  
 2608 AACTGGGGCTGGTCCCTTTGGGGTGGACGTAGATTAAAGCCCTTAATACAGATCCTG 2667  
 861 AsnTyrGlyCysGlyAlaPheGlyGlyAspAlaArgLeuLysAlaLeuLeuGlnIleLeu 880  
 2668 GCAGCTGCTGGCTGAACGTGACGTGGTGTATTATTCACCTTTGGGACTCAGAGTTGATG 2727  
 881 AlaAlaAlaAlaAlaGluArgAspValValPheThrPheGlyAspSerGluLeuMet 900  
 2728 AGACACATTACAGCATGCACATCTTCTTACCAGAGAGAGTGGATGTTGGAAAGTG 2787  
 901 ArgAspIleIysSerMetHisThrPheLeuThrGluArgLysLeuAspValGlyLysVal 920  
 2788 TACAAGTTATTGCTAGATACATACTCAATGAAGAATGCAAACTGTTCCACCTCGACCA 2847  
 921 TyrLysLeuLeuLeuArgTyrTyrAsnGluGluCysArgAsnCysSerThrProGlyPro 940  
 2848 GACATCAAGCTTATCCATTTCATATACCATGCTTTCAGTCAAGTGCAGAGACCACTGAC 2907  
 941 AspIleLysLeuTyrProPheIleTyrHisAlaValGluSerSerAlaGluThrThrAsp 960  
 2908 ATGCCAGCAGACAGGAGGAGGCC 2931  
 961 MetProGlyGlnLysAlaGlyThr 968

RESULT 4  
 ABG72280  
 ID ABG72280 standard; protein; 968 AA.  
 XX AC ABG72280;  
 XX DT 13-MAR-2003 (first entry)  
 XX DE Murine poly (ADP-ribose) glycohydrolase (PARG) enzyme.  
 XX KW Murine; poly (ADP-ribose) glycohydrolase; PARG; PARG expression;  
 KW cellular response; DNA damage; neoplastic disorder inducing agent;  
 KW oxidative stress; neoplastic disorder; myocardial infarction;  
 KW vascular stroke; neurodegenerative disorder; Alzheimer's disease;  
 KW parkinson's disease; Huntington's disease; inborn genetic error;  
 KW reperfusion; ischaemia; aging; neurotoxicity; PARG activity; mouse;  
 KW cytosolic; neuroprotective; motropic; antiparkinsonian; cardiac;  
 KW vasotropic; anticonvulsant; cerebroprotective; enzyme.  
 XX OS Mus musculus.  
 XX PN US2002132328-A1.  
 XX PD 19-SEP-2002.  
 XX PF 09-OCT-2001; 2001US-00973451.  
 XX PR 01-MAY-1998; 98US-0083768P.  
 XX PR 30-APR-1999; 99US-00302812.  
 XX PA (JACO/) JACOBSON M K.  
 XX PA (JACO/) JACOBSON E L.  
 XX PA (AMEJ/) AME J.  
 XX PA (LINW/) LIN W.  
 XX PI Jacobson MK, Jacobson EL, Ame J, Lin W;  
 XX WPI; 2003-155895/15.  
 XX DR N-PSDB; ABX14479.  
 XX XX New nucleic acid molecule encoding a polypeptide with poly (ADP-ribose)  
 PT glycohydrolase activity, for preventing, treating, or ameliorating a  
 PT disease condition, e.g. neoplastic disorder, myocardial infarction or  
 PT vascular stroke.  
 XX PS Claim 28; Fig 16; 86pp; English.

The present invention relates to the isolation of poly (ADP-ribose) glycohydrolase (PARG) from several species, and the polynucleotide sequences encoding them. Methods for inhibiting PARG expression or overexpressing PARG are also disclosed. PARG is involved in the cellular response to DNA damage, and is associated with the body's response to neoplastic disorder inducing agents and oxidative stress. The polynucleotide sequences encoding PARG and PARG modulators are useful for preventing, treating, or ameliorating diseases such as neoplastic disorders, myocardial infarction, vascular stroke, neurodegenerative disorders (e.g. Alzheimer's disease, Parkinson's disease or Huntington's disease), inborn genetic errors, reperfusion following ischaemia, aging, and neurotoxicity. The polynucleotide sequences are also useful in gene therapy. The methods are useful for identifying an agent that modulates PARG activity, identifying a mutant PARG allele in an individual, or screening candidate molecules for PARG modulating activity. The present sequence represents murine PARG enzyme

XX Sequence 968 AA;  
 SQ

Alignment Scores:  
 Pred. No.: 0 Length: 968  
 Score: 5164.00 Matches: 968  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 76.24% Indels: 0  
 DB: 6 Gaps: 0

US-09-302-812-5 (1-3814) x ABG72280 (1-968)

QY 28 ATGAGTCGGGCCCCGGCTGGAGCCCTGCAGCAAGCGCGCTGGCGCGCTGGAAC 87  
 DB 1 MetSerAlaGlyProGlyTyrGluProCysThrLysAlaArgTyrGlyAlaGlyThr 20  
 QY 88 TCTGGCGGAGTCCCTCGACTCCCGGAGTCTCCCTGGCAGGAGAGCGCTGTTCTCGAC 147  
 DB 21 SerAlaProThrAlaSerAspSerArgSerPheProGlyArgGlnArgValLeuAsp 40  
 QY 148 CCCAAGGACGCTCCCGTCCAGATTCAGGGTCCCTCCGCTCCGCGAGCCCTGCTCGGG 207  
 DB 41 ProllysAspAlaProValGlnPheArgValProProSerSerProAlaCysValSerGly 60  
 QY 208 CGGCGGGACCGCAGACAGGCAAGCCACCTCGTGTGTTTCAAAACAAAGACTATTACT 267  
 DB 61 ArgAlaGlyProHisArgGlyAsnAlaThrSerPheValPheLysGlnLysThrIleThr 80  
 QY 268 ACTTGGATGGATACTAAAGGACCCCAAGACAGCTGAATCAGAAAGTAAAGAAACAACAT 327  
 DB 81 ThrTyrMetAspThrLysGlyProLysThrAlaGluSerGluSerLysGluAsnAsn 100  
 QY 328 ACAAGAATTGACTCCATGATGAGTTCTGTGAGAAAGATTAATTTTACCCACATAAGGTG 387  
 DB 101 ThrArgIleAspSerMetMetSerSerValGlnLysAspAsnPheTyrProHisLysVal 120  
 QY 388 GAAATTTGGAATGTTCTCAGCTAAATCTTGATAATCACCACACAGAAAGAGTTCA 447  
 DB 121 GluLysLeuGluAsnValProGlnLeuAsnLeuAspLysSerProThrGluLysSerSer 140  
 QY 448 CAGTATTTTGAACCAACAGCAGACTCGGAGTGTGTCAAGTGGCAGAAAGAGGAGCAT 507  
 DB 141 GlnTyrLeuAsnGlnGlnThrAlaSerValCysLysTyrGlnAsnGluGlyLysHis 160  
 QY 508 GCAGAACAGCTTTTGGGAAAGTGGAGCTCCCGGGGAGTCCCGCTACCAAGCAGCTTACT 567  
 DB 161 AlaGluGlnLeuLeuAlaSerGluProProAlaGlyThrProLeuProLysGlnLeuSer 180  
 QY 568 AATGCTAATTTGGTTCAGTCACTCCACACATGATGACACAGTGCACAGATCATCAAGAA 627  
 DB 181 AsnAlaAsnIleGlyGlnSerProHisThrAspAspHisSerAspThrAspHisGluGlu 200  
 QY 628 GACAGAGACAATCAGCAGTCTTCTTACACCTATAAACTTGCATAAACAAGCAACAGTA 687  
 DB 201 AspArgAspAsnGlnGlnPheLeuThrProLysLeuAlaAsnThrLysProThrVal 220

QY 688 GGAGATGGCGAGCGCAGAGCAACTGTAGTGCAGTGCAGTCTGCCAGTCTGTGAAGAC 747  
DB 221 GlyAspGlyGlnAlaArgSerAsnCysCysSerGlySerArgGlnSerValIysAsp 240  
QY 748 TGTACAGCTGTCAACAGGAGGAGGTGATGCTTACAGAGAGTCCCTTGTGTGACAGATT 807  
DB 241 CysThrGlyCysGlnGlnGluValAspValIeuProGluSerProIeuSerAspVal 260  
QY 808 GGTCCGAGGACATTGGAACTGGACCAACAAAATGACAACTTGAACCTGGACAGAAAGC 867  
DB 261 GlyAlaGluAspIleGlyThrGlyProIysAsnAspAsnIysLeuThrGlyGlnGluSer 280  
QY 868 AGCCTAGTGTATTCGCTCCATTTGAGAAAGAAAGTGCAGTGCAGTGCAGTGCAGTGA 927  
DB 281 SerIeuGlyAspSerProPheGluIysGluSerGluProGluSerProMetAspVal 300  
QY 928 GACAACTCAGAAACAGTTGTCAAGATTCAGAACGAGATGAAGAAACAAGTCCAGTCTTT 987  
DB 301 AspAsnSerArgAsnSerCysGlnAspSerGluAlaAspGluGluThrSerProValPhe 320  
QY 988 GATGAGCAAGATGATCGTTCTCCCAACAGCAATAAATTTCAAGTTGCCAAGCAAGA 1047  
DB 321 AspGluGlnAspAspArgSerSerGlnThrAlaAsnIysLeuSerSerCysGlnAlaArg 340  
QY 1048 GAACTGTATGCGCATTTAGAAACGGTATTGTACTAAGGAAAGTGAAGTTAGATTGCAT 1107  
DB 341 GluAlaAspGlyAspLeuArgIysArgTyrIeuThrIysGlySerGluValArgIeuHis 360  
QY 1108 TTCCAATTTGAGGAGAAATAATCTCGGACAGTCACTTAATGCCAAGCCATCTCGA 1167  
DB 361 PheGlnPheGluGlyGluAsnAlaGlyThrSerAspLeuAlaIysProSerGly 380  
QY 1168 AACTCTTCTAGCTTAATGTAGAGTGTAGAATTTCCACAGCAGCATGGAAAAAGGAAATCT 1227  
DB 381 AsnSerSerLeuAsnValGluCysArgSerSerIysGlnHisGlyIysArgAspSer 400  
QY 1228 AAAATTAAGATCATTTGATGAAATTTCCAGTGCAGAGGACAGAAAGAAACAATGT 1287  
DB 401 LysIleThrAspHisPheMetArgIleSerIysSerGluAspArgArgIysGluGlnCys 420  
QY 1288 GAAGTCAGACATCAAGACAGAAAGGAGATTTCCAAATATACATCCCACTTAACCTCCCT 1347  
DB 421 GluValArgHisGlnArgThrGluArgIysIleProIysTyrIleProAsnLeuPro 440  
QY 1348 CCAGAGAAAGTGGCTGGGAACTCCTATTGAGGAAATGAGAAAAATCGCTCGGTGGG 1407  
DB 441 ProGluIysIysTripleuGlyThrProIleGluMetArgIysMetProArgCysGly 460  
QY 1408 ATCCATTTGCTTCTTAAGACCATCTGCAAGTGCACAGTCACTGTTCGGGTAGACCTT 1467  
DB 461 IleHisLeuProSerLeuArgProSerAlaSerHisThrValThrValArgValAspLeu 480  
QY 1468 CTGAGCAGGAGAGGTTCCGAAACCTTTTCCAAACACATTACAAAGATTGTGGGATAAC 1527  
DB 481 LeuArgAlaGlyGluValProIysProPheProThrHisTyrIysAspLeuTriAspAsn 500  
QY 1528 AAACATGTCAAAATGCTTTGTCGGAACAAAATTTGTACCCCTGTGGAGATGAGATGTGT 1587  
DB 501 LysHisValIysMetProCysSerGluGlnAsnLeuTyrProValGluAspGluAsnGly 520  
QY 1588 GAGCGAACCTGCGGGAGTGTGGGAGCTTCATTCAGACTGCACCTTCTCAACAATTCACA 1647  
DB 521 GluArgThrAlaGlySerArgTyrGluLeuIleGlnThrAlaLeuLeuAsnIysPheThr 540  
QY 1648 CGACCCCAAGACTTGAAGATGCGATTCTGAATACAAATGTGGCAATTTCTAAGAAATGG 1707  
DB 541 ArgProGlnAsnLeuIysAspAlaIleLeuIysTyrAsnValAlaTyrSerIysIysTriP 560  
QY 1708 GACTTTACAGCTTTGTGTTGATTTCTGGGATAAGGTACTTTGAAGACGACGAGGCCCAACAT 1767  
DB 561 AspPheThrAlaLeuValAspPheTyrAspIysValIeuGluGluAlaGluAlaGlnHis 580  
QY 1768 TTATATCAGTCCATTTTACCTGACATGGTGAATAATTCGACTCTGTCTGCCAATATTTCG 1827

DB 581 LeuTyrGlnSerIleLeuProAspMetValIysIleAlaLeuCysLeuProAsnIleCys 600  
QY 1828 ACCAGCCATACCACTCTCTGAAACAGAGATGAATCAITCTGTCTCAGATGTCAAGAA 1887  
DB 601 ThrGlnProIleProLeuLeuIysGlnIysMetAsnHisSerValThrMetSerGlnGlu 620  
QY 1888 CAGATCGCCAGTCTTTTAGCTAATGCTTTCTTCTGCACATTTCCCGACCGAATGCCAAG 1947  
DB 621 GlnIleAlaSerLeuLeuAlaAsnAlaPhePheCysThrPheProArgArgAsnAlaIys 640  
QY 1948 ATGAAATCGCAGTATTCTAGTTACCCACAGACATTAACTCAATCGTGTGTGTGAAGACGT 2007  
DB 641 MetIysSerGluTyrSerSerTyrProAspIleAsnPheAsnArgLeuPheGluGlyArg 660  
QY 2008 TCATCAAGGAAACCCAGAAACCACTCTTCTGCTACTTTCGAAAGAGTACACAG 2067  
DB 661 SerSerArgIysProGluIysLeuIysThrLeuPheCysTyrPheArgArgValThrGlu 680  
QY 2068 AAAAACTACAGATTGGTGACATTTACAGACAGAGTCTTGAAGATTTTCCAGAATGG 2127  
DB 681 LysIysProThrGlyLeuValThrPheThrArgGlnSerLeuGluAspPheProGluTriP 700  
QY 2128 GAAAGGTGTGAAAACGCTCTGCACACGCTTACACGCTTACCTTACGAGGGTACCATGAAGGC 2187  
DB 701 GluArgCysGluIysProIeuThrArgLeuHisValThrTyrGluGlyThrIleGluGly 720  
QY 2188 AACCGCCGAGCATGCTACAGTGGATTTGCAAAATCGTTTGTGGAGGTGGTGCAT 2247  
DB 721 AsnGlyArgGlyMetLeuGlnValAspPheAlaAsnArgPheValGlyGlyValThr 740  
QY 2248 GGTCCGGACTGTACAGAGAAATCAGATTTTAACTCAATCCTGAATTCATGTTTCA 2307  
DB 741 GlyAlaGlyLeuValGlnGluGluIleArgPheLeuIleAsnProGluLeuIleValSer 760  
QY 2308 CGGCTGTTTCACTGAGGTGCTGGATCACAATGAGTCTCTTATTATCAGAGTACTGAACAG 2367  
DB 761 ArgLeuPheThrGluValLeuAspHisAsnGluCysLeuIleIleThrGlyThrGluGln 780  
QY 2368 TACAGTGAATACAGAGCTATGCTGAAACTTATGCTGGCGCCGAGCAGCAGTGAACAGCT 2427  
DB 781 TyrSerGluTyrThrGlyTyrAlaGluThrTyrArgTyrAlaArgSerHisGluAspGly 800  
QY 2428 AGTGAAGAGCAGATTGGCAGCGCTGCAGGAGATCGTTCATTCAGCGCACTTCCAC 2487  
DB 801 SerGluIysAspAspTriPleuArgCysThrGluIleValAlaIleAspAlaLeuHis 820  
QY 2488 TTCAGACGCTACCTCGATCAGTTTGTGCTCAGAAAGTGAGACGTGAGCTTAACAGGCT 2547  
DB 821 PheArgArgTyrLeuAspGlnPheValProGluIysValArgGluLeuAsnIysAla 840  
QY 2548 TACTCGGATTCCTCCGCTCGGAGTTCCTTCTGCAAAATCTTCTGCAGTGGCCACGGA 2607  
DB 841 TyrCysGlyPheLeuArgProGlyValProSerGluAsnLeuSerAlaValAlaThrGly 860  
QY 2608 AACTCGGCTGTGGTGGCTTTGGGGTGCAGCTAGATTAAAGCCCTTAATACAGATCCTG 2667  
DB 861 AsnTriPleuGlyCysGlyAlaPheGlyGlyAspAlaArgLeuIysAlaLeuIleGlnIleLeu 880  
QY 2668 GCAGCTGTGCGGTGAAACGTGACGTGTTTATTTTTCACCTTTGGGACTCAGATTGATG 2727  
DB 881 AlaAlaAlaAlaGluArgAspValValTyrPheThrPheGlyAspSerGluLeuMet 900  
QY 2728 AGAGCATTTACAGATCGACATTTCTTCTCCAGAGAGAGTCTGATGTTGGAAAAGTG 2787  
DB 901 ArgAspIleTyrSerMetHisThrPheLeuThrGluArgIysLeuAspValGlyIysVal 920  
QY 2788 TACAAGTTATTCTTGTAGTACTTACAAATGAAGAAATGCAGAAATCTGTTCACCCCTCGACA 2847  
DB 921 TyrIysLeuLeuLeuArgTyrTyrAsnGluGluCysArgAsnCysSerThrProGlyPro 940  
QY 2848 GACATCAAGCTTTATCCATTCATATACCATGCTGTGTGAGTCAAGTGCAGACCACTGAC 2907



1084 AAGGAAGTGAAGTGTAGATTTCATTTCCAAATTTGAA---GGAGAAAATAATGCTGGGACC 1140  
 Db  
 360 LysGlyGlyGluValArgLeuHisPheGlnPheGluGlyGlyGluSerArgThrGlyMet 379  
 QY  
 1141 AGTGACTTAATGCAAGCCATCTGGAACTCTTTCTAGCCCTTAATGTAGAGTGTAGAGT 1200  
 Db  
 380 AsnAspLeuAsnAlaLysLeuProGlyAsnIleSerSerLeuAsnValGluCysArgAsn 399  
 QY  
 1201 TCCAGAGCATGGAAGAGGATCTTAATATTACAGATCATTTTCATGAGATTTCCAG 1260  
 Db  
 400 SerLysGlnHisGlyLysAspSerLysIleThrAspHisLeuMetArgLeuProLys 419  
 QY  
 1261 TCAGAGGACAGAGAAAGAACAAATGCAAGTCAGACATCAAGAACAGAGAGAGATTT 1320  
 Db  
 420 AlaGluAspArgGlyGlyGlnTrpGluThrLysHisGlnArgThrGluArgLysIle 439  
 QY  
 1321 CCAAAATACATCCCACTCACTCTCCAGAGAGAGAGTGGCTGGAACTCTTATGAG 1380  
 Db  
 440 ProLysTyrValProProHisLeuSerProAspLysLysTrpLeuGlyThrProIleGlu 459  
 QY  
 1381 GAATGAGAAATGCTCGGTGGATCCATTTGCTTCCCTTAAGACCATCTGCAAGT 1440  
 Db  
 460 GluMetArgMetProArgCysGlyLeuArgLeuProLeuLeuArgProSerAlaAsn 479  
 QY  
 1441 CACAGAGTGAAGTTCGGGTAGACCTTCTGAGAGAGAGAGTTCGGAACCTTTTCCA 1500  
 Db  
 480 HisThrValThrIleArgValAspLeuLeuArgAlaGlyGluValProLysProPhePro 499  
 QY  
 1501 ACACATTACAAAGATTTGTGGATACCAACATGTGAATAATCCCTTTGTCGGAACAAAC 1560  
 Db  
 500 ThrHisTyrLysAspLeuTrpAspAsnLysHisValLysMetProCysSerGluGlnAsn 519  
 QY  
 1561 TTGTACCTCTGGAAGATCAGATGCTGAGCGAACTGACGGAGTAGTGGGAGCTCAT 1620  
 Db  
 520 LeuTyrProValGluAspGluAsnGlyGluArgThrAlaGlySerArgTrpGluLeuIle 539  
 QY  
 1621 CAGACTGCACTTCTCAACAAATTCACAGCCCAAGAACTTGAAGGATCGGATTTCTGAAA 1680  
 Db  
 540 GlnThrAlaLeuLeuAsnLysPheThrArgProGlnAsnLeuLysAspAlaIleLys 559  
 QY  
 1681 TACAATGTGGCATATTCTAAGAAATGGGACTTTACAGCTTTGGTTGATTTCTGGGATAG 1740  
 Db  
 560 TyrAsnValAlaTyrSerLysLysTrpAspPheThrAlaLeuIleAspPheTrpAspLys 579  
 QY  
 1741 GTACTTGAGAGACAGAGCCCAACATTTATATCATCTCCTATTTACTCAGATGTTGAAA 1800  
 Db  
 580 ValLeuGluGluAlaGluAlaGlnHisLeuTyrGlnSerIleLeuProAspMetValLys 599  
 QY  
 1801 ATTGCACTCTGTCTGCAAAATATTTCAGCCCAATACCACTCTCTGAAACAGAGATG 1860  
 Db  
 600 IleAlaLeuCysLeuProAsnIleCysThrGlnProIleProLeuLysGlnLysMet 619  
 QY  
 1861 AATCATCTGTGCAAGTGTACAGGAACAGATGCCAGCTTTTACTGATGCTTTCTTC 1920  
 Db  
 620 AsnHisSerIleThrMetSerGlnGluGlnIleAlaSerLeuLeuAlaAsnAlaPhePhe 639  
 QY  
 1921 TGCACATTTCCCGGACGAGATGCAAGATCGAATCGAGTATTCTAGTTTACCCAGACATT 1980  
 Db  
 640 CysThrPheProArgAsnAlaLysMetLysSerGluTyrSerSerTyrProAspIle 659  
 QY  
 1981 AACTTCAATCGGTGTTTGAAGGACGTTCTATCAAGGAAACAGAGAAACCTGAAACACTC 2040  
 Db  
 660 AsnPheAsnArgLeuPheGluGlyArgSerSerArgLysProGluLysLeuLysThrLeu 679  
 QY  
 2041 TTCTGTACTTTGCAAGATTCACAGAGAAACCTACAGATGTTGATGATTTTACAGA 2100  
 Db  
 680 PheCysTyrPheArgArgValThrGluLysLysProThrGlyLeuValThrPheThrArg 699  
 QY  
 2101 CAGAGCTTCAAGATTTTCCAGATGGAAGGTGTGAAAGCTCTGACACGCTTACAC 2160  
 Db  
 700 GlnSerLeuGluAspPheProGluTrpGluArgCysGlnLysProLeuThrArgLeuHis 719  
 QY  
 2161 GTCACTTACGAGGCTACCATAGAGGCAACGGCCGAGGCATGCTACAGGTGGATTTTGCA 2220

720 ValThrTyrGluGlyThrIleGluGluAsnGlyGlnGlyMetLeuGlnValAspPheAla 739  
 QY  
 2221 AATCGTTTTGTGAGGTGTGTGACTGTGCGGACTTGTACAAAGAAAGAAATCAGATTT 2280  
 Db  
 740 AsnArgPheValGlyGlyGlyValThrSerAlaGlyLeuValGlnGluIleArgPhe 759  
 QY  
 2281 TTAATCAATCCTGAATTTGTTTTCACGCTGTCTCAGGTCGTGGATCACAATGAG 2340  
 Db  
 760 LeuIleAsnProGluLeuIleIleSerArgLeuPheThrGluValLeuAspHisAsnGlu 779  
 QY  
 2341 TGTCTTATTATCACAGTACTGAACACTGACAGTAAATACACAGGCTATCTGTAACCTTAT 2400  
 Db  
 780 CysLeuIleIleThrGlyThrGluGlnTyrSerGluTyrThrGlyTyrAlaGluThrTyr 799  
 QY  
 2401 CGTTGGCCCGCAAGCCATGAAGTGGAGTGAAGAGGACGATTGCGACGCGCTGCAAG 2460  
 Db  
 800 ArgTrpSerArgSerHisGluAspGlySerGluArgAspAspCysGluArgArgCysThr 819  
 QY  
 2461 GAGATCGTTGCCATTGACGCACTTCACTTCAGACGCTACCTCGATCAGTTTGTGCTGAG 2520  
 Db  
 820 GluIleValAlaIleAspAlaLeuHisPheArgArgTyrLeuAspGlnPheValProGlu 839  
 QY  
 2521 AAGTGAGACGTGAGCTTAAACAGGCTTACTGCGGATTCCTCCGTCTGGAAGTTCCTTCT 2580  
 Db  
 840 LysMetArgArgGluLeuAsnLysAlaTyrCysGlyPheLeuArgProGlyValSerSer 859  
 QY  
 2581 GAAATCTTTCTGAGTGGCCAGCGGAACTTGGGCTGTGCTGCTTTGGGGGTGACGCT 2640  
 Db  
 860 GluAsnLeuSerAlaValAlaThrGlyAsnTrpGlyCysGlyAlaPheGlyGlyAspAla 879  
 QY  
 2641 AGATTAAGCTTAAATACAGATCTGCGACGCTGCGGCTGACGTAACGTAACGTTTAT 2700  
 Db  
 880 ArgLeuLysAlaLeuIleGlnIleLeuAlaAlaAlaGluArgAspValValTyr 899  
 QY  
 2701 TTCACCTTTGGGACTCAGAGTGTGATGAGACATTTTACAGCATGCACACTTCTCTTACC 2760  
 Db  
 900 PheThrPheGlyAspSerSerGluLeuMetArgAspIleTyrSerMetHisIlePheLeuThr 919  
 QY  
 2761 GAGAGGAAGCTGGATGTTTGAAGAGTGTACAGTTATTGCTTAGTACTACATCAAGAA 2820  
 Db  
 920 GluArgLysLeuThrValGlyAspValTyrLysLeuLeuLeuArgTyrTyrAsnGluGlu 939  
 QY  
 2821 TGCAGAAAATGTTTCCACCCCTGGACAGACATCAAGCTTTTATCCATTCATATACCATGCT 2880  
 Db  
 940 CysArgAsnCysSerThrProGlyProAspIleLysLeuTyrProPheIleTyrHisAla 959  
 QY  
 2881 GTTGAGTCAAGTGCAGAGACCATGACATGCCAGACAGAGAGGCGAGGACCC 2931  
 Db  
 960 ValGluSerCysAlaGluThrAlaAspHisSerGlyGlnArgThrGlyThr 976

RESULT 6  
 AAU76021  
 ID AAU76021 standard; protein; 976 AA.  
 XX  
 AC AAU76021;  
 XX  
 DT 08-MAY-2002 (first entry)  
 XX  
 DE Human poly (ADP-ribose) glycohydrolase (PARG) protein sequence.  
 XX  
 KW Human; poly (ADP-ribose) glycohydrolase; PARG; DNA repair; ADP-ribose;  
 KW adenosine diphosphate-ribose; neoplastic disorder; adenocarcinoma;  
 KW leukaemia; lymphoma; melanoma; sarcoma; teratocarcinoma;  
 KW hyperplasia; hypertrophy; reperfusion; ischaemia; heart attack; stroke;  
 KW neurodegenerative disease; neurological disorder; Alzheimer's disease;  
 KW Huntington's disease; Parkinson's disease.  
 XX  
 OS Homo sapiens.  
 XX  
 PN US6337202-B1.  
 XX  
 XX 08-JAN-2002.  
 PD



XX PF 23-FEB-2000; 2000US-00511477.  
XX PR 01-MAY-1998; 98US-0083768P.  
XX PR 30-APR-1999; 99US-00302812.  
XX (KENT ) UNIV KENTUCKY RES FOUND.  
XX Jacobson MK, Jacobson EL, Ame J, Lin W;  
XX WPI; 2002-163240/21.  
XX N-PSDB; ABK14932.  
XX Novel isolated and purified poly (ADP-ribose) glycohydrolase protein  
XX PT which catalyses release of ADP-ribose from ADP ribose polymer, useful for  
XX PT treating neoplastic and neurological disorders, heart attack and stroke.  
XX PS Claim 2; Col 55-60; 81pp; English.  
XX The present invention relates to a new poly (ADP-ribose) glycohydrolase  
XX CC (PARG) protein which catalyses release of ADP-ribose from an ADP  
XX CC (adenosine diphosphate)-ribose polymer. The PARG molecule of the  
XX CC invention is useful for generating antibodies and can be inhibited or  
XX CC activated for diagnosing and treating neoplastic disorders such as  
XX CC adenocarcinoma, leukaemia, lymphoma, melanoma, myeloma, sarcoma,  
XX CC teratocarcinoma, hyperplasia and hypertrophy, reperfusion following  
XX CC ischaemia, heart attack, stroke, neurodegenerative diseases, neurological  
XX CC disorders including Alzheimer's, Huntington's and Parkinson's diseases,  
XX CC and related conditions. PARG levels may be enhanced to suppress DNA  
XX CC repair and increase the cell's susceptibility to chemotherapy drugs.  
XX CC Antagonists of PARG are administered to treat or prevent neoplastic  
XX CC disorders. The present amino acid sequence represents the human PARG  
XX CC protein of the invention. This protein is one of several PARG proteins  
XX CC (AAU76020-AAU76024) of the invention  
XX SQ Sequence 976 AA;

Alignment Scores:  
Pred. No.: 0 Length: 976  
Score: 4395.00 Matches: 837  
Percent Similarity: 90.99% Conservative: 52  
Best Local Similarity: 85.67% Mismatches: 78  
Query Match: 64.89% Indels: 10  
DB: Gaps: 6

US-09-302-812-5 (1-3814) x AAU76021 (1-976)

QY 28 ATGAGTGGGGCCCCGGCTGGAGCCCTGCACGAAGCG---CGCTGGGGCGCGCTGGA 84  
Db 1 MetAsnAlaGlyProGlyCysGluProCysThrLysAlaThrArgTrpGlyAlaAa--- 19  
QY 85 ACTTCTGCGCGGACTGCTCGGACTCCGAGCTCTCCCTGGCAGGAGAGCGCGTGTCTC 144  
Db 20 ThrThrSerProAlaAlaSerAspAlaArgSerPheProSerArgGlnArgValLeu 39  
QY 145 GACCCCAAGACGCTCCCGTCCAGTTCAGGTCCTCCCTCCGCTCCGCGAGCTCGCTCG 204  
Db 40 AspProLysAspAlaHisValGlnPheArgValProProSerSerProAlaCysValPro 59  
QY 205 GGGCGGGCGGAGCGGACAGAGGACCGACCTCGTCTCTTTTCAACAAAGACTATT 264  
Db 60 GlyGlnAlaGlyGlnHisArgGlySerAlaThrSerLeuValPheLysGlnLysThrile 79  
QY 265 ACTACTTGGATGGATCTAAAGGACCAACAGACGCTGAATCAGAA-----AGTAAA 315  
Db 80 ThrSerTrpMetAspThrLysGlyLysThrLagLysGluSerLeuAspSerLys 99  
QY 316 GAAACACAAATACAGAATTTGACTCCATGATGATGTTCTGTGCAAGAAGATACTTTTAC 375  
Db 100 GluAsnAsnAnThrArgIleGluSerMetMetSerSerValGlnLysAspAsnPheTyr 119  
QY 376 CCACATAGGTGGAAAAATTGAAAAATTTCTCCTCAGCTAAATCTTGATAAATCACCCACA 435

Db 120 GlnHisAsnValGluLysLeuValAsnValSerGlnLeuSerLeuAspLysSerLeuThr 139  
QY 436 GAAAAGAGTTCACAGTATTGTGAACCAACAGACAGACTGCGAGTGTGTGCAAGTGGCAAT 495  
Db 140 GluLysSerThrGlnTyrLeuAsnGlnHisGlnThrAlaAlaMetCysLysTrpGlnAsn 159  
QY 496 GAAGGGAAGCATGCAGAACGCTTTTGGCAAGTGAAGCTCCCGCGGGGACTCCGTACCA 555  
Db 160 GluGlyLysHisThrGluGlnLeuLeuGluSerGluProGlnThrValThrLeuValPro 179  
QY 556 AAGCAGCTTAGTAATGCTAACTAGTTCAGTCACTCCCACTGATGATGACACAGTGCACA 615  
Db 180 GlnGlnPheSerAsnAlaAsnLysAspArgSerProGlnAsnAspAspHisSerAspThr 199  
QY 616 GATCATGAAGAAGACAGACAGCAATCAGAGTTCCTTACACCTATTAACCTGCAATACA 675  
Db 200 AspSerGluGluAsnArgAspAsnGlnGlnPheLeuThrValLysLeuAlaAsnAla 219  
QY 676 AAGCCCAACAGTAGGAGATGGGCGAGCCAGA-----AGCAACTGTAAAGTGCAGTGA 726  
Db 220 LysGlnThrThrGluAspGluHisAlaArgGluAlaLysSerHisGlnLysCysSerLys 239  
QY 727 TCTCGCCAGTCTGTGAAGACTGTACAGCTGTCAACAGAGGAGGAGTGTGCTACCA 786  
Db 240 SerCysHisProGlyGluAspCysAlaSerCysGlnAspGluLysLeuAspValPro 259  
QY 787 GAGAGTCTTTTTCAGATTTGGTCCGAGACATTGGAACCTGGACCAAAATGCAAC 846  
Db 260 LysSerProLeuSerAspValGlySerGluAspValGlyThrGlySerLysAsnAspAsn 279  
QY 847 AAATGACTGGACAGACAGCTAGTGTGATTCCTCCCTCCATTTGAGAAAGAAAGTGAG 906  
Db 280 LysLeuLysArgGlnGlnSerCysLeuGlyAsnSerProProPheGlnLysGluSerGlu 299  
QY 907 CCTGAGTCAACATGATGTAGCAACTGCAACCTGAGAAACAGTTGTCAAGATTCAGAACGAG 966  
Db 300 ProGluSerProMetAspValAspAsnSerLysAsnSerCysGlnAspSerGluAlaAsp 319  
QY 967 GAAGAAACAGTCCAGTCTTTGATGAGCAAGATGAT---CGTTCTCCCAAAACAGCAAT 1023  
Db 320 GluGluThrSerProGlyPheAspGluGlnGluAspGlySerSerGlnThrAlaAsn 339  
QY 1024 AATCTTCAAGTTGCGCAACAGAGAGCTGTGGGCTCTTAGGAAACCGTATTTGACT 1083  
Db 340 LysProSerArgPheGlnAlaArgAspAlaAspIleGluPheArgLysArgTyrSerThr 359  
QY 1084 AAGGGAAGTGAAGTATGATTTCCATTTCCAAATTTGAA---GGAGAAATATATGCTGGACC 1140  
Db 360 LysGlyGlyGluValArgLeuHisPheGlnPheGluGlyGlyGluSerArgThrGlyMet 379  
QY 1141 AGTGACTTAATGCCAAGCCATCTGAAACTCTTCTAGCTTAAATGATGAGTGTAGAGT 1200  
Db 380 AsnAspLeuAsnAlaLysLeuProGlyAsnLysSerSerLeuAsnValGluCysArgAsn 399  
QY 1201 TCCACACAGCATCGAAAAAGGATCTTAAATATACAGATCATTTTCATGAAATTTCCAG 1260  
Db 400 SerLysGlnHisGlyLysAspSerLysLysLysLysLysLysLysLysLysLysLys 419  
QY 1261 TCAGAGCAGACAGAAAAAGAACAACTGTGAAGTCAGACATCAAAAGACAGAAAGAGAGATT 1320  
Db 420 AlaGluAspArgArgLysGluGlnTrpGluThrLysHisGlnArgThrGluArgLysLys 439  
QY 1321 CAAAATACATCCCACTTAACCTCCCTCCAGAGAAAGAGTGGCGGAACTCCCTATTGAG 1380  
Db 440 ProLysTyrValProProHisLeuSerProAspLysLysLysLysLysLysLysLysLys 459  
QY 1381 GAAATCAGAAAAATGCTCGCTGGGATCCATTTGCTTCTTCTTAAACCATCTGCAAGT 1440  
Db 460 GluMetArgArgMetProArgCysGlyLysLysLysLysLysLysLysLysLysLysLys 479  
QY 1441 CACACAGTCACTGTTGGGTAGACCTTCTGAGACGAGGAGGTTCCGAAACCTTTTCCA 1500  
Db 480 HisThrValThrIleArgValAspLeuLeuAlaGlyGluValProLysProPhePro 499



1501 ACACATTACAAAGATTGTGGATACAAACATGTGAAATGCTTCTTTCGGAAACAAAC 1560  
 500 ThrHisTyrIysAspLeuTrpAspAsnLysHisValIysMetProCysSerGluGlnAsn 519  
 1561 TTGTCACCTCTGGAGATGAAATGGTGCAGGAACTGCAGGAGCTAGGTGGAGCTCAT 1620  
 520 LeuTyrProValGluAspGluAsnGlyGluArgThrAlaGlySerArgTrpGluLeu 539  
 1621 CAGACTGCACCTCTCAACAAATTCACACAGCCAGCACTTGAAGGATGCAATTCGAAA 1680  
 540 GlnThrAlaLeuLeuAsnLysPheThrArgProGlnAsnLeuLysAspAlaLeuLys 559  
 1681 TACAATGTGGCAATTCAGAAATGGACCTTACAGCTTGGTGTGATTTCTGGATAAG 1740  
 560 TyrAsnValAlaTyrSerLysLeuTrpAspPheThrAlaLeuLysAspPheTrpAspLys 579  
 1741 GTACTTGAAGAGCAGAGCCCAACATTTATCATGCTCCATTTTACCTGACATCGTGA 1800  
 580 ValLeuGluGluAlaGluAlaGlnHisLeuTyrGlnSerIleLeuProAspMetValLys 599  
 1801 ATTGCACCTCTCTGCCAAATATTTCACCCAGCAATACACCTCTGAAACAGAGATG 1860  
 600 IleAlaLeuCysLeuProAsnIleCysThrGlnProIleProLeuLeuLysGlnLysMet 619  
 1861 AATCATCTCTCAGATGTCACAGAACAGATCGCGAGCTTTTAGCTAATGCTTCTTC 1920  
 620 AsnHisSerIleThrMetSerGlnGluGlnIleAlaSerLeuLeuAlaAsnAlaPhePhe 639  
 1921 TGCATTTTCCCGCAGCAATGCGAAGTGAATCGGAGTATCTAGTTACCCAGACATT 1980  
 640 CysThrPheProArgAsnAlaLysMetLysSerGluTyrSerSerTyrProAspIle 659  
 1981 AACTTCATCGGTGTTTGAAGGAGCTTCATCAAGAAACAGAAACAACTGAAACACTC 2040  
 660 AsnPheAsnArgLeuPheGluGlyArgSerSerArgLysProGluLysLeuLysThrLeu 679  
 2041 TTCTGCTATTTCGAAGAGTCACAGAGAAACAACTCAGATTTGCTGACATTTACAGA 2100  
 680 PheCysTyrPheArgArgValThrGluLysLysProThrGlyLeuValThrPheThrArg 699  
 2101 CAGAGTCTGAGATTTTCCAGATGCGAAGGCTGAAAGCTCTGACACGCTTACAC 2160  
 700 GlnSerLeuGluAspPheProGluTrpGluArgCysGluLysProLeuThrArgLeuHis 719  
 2161 GTCACCTACGAGGTACCAAGAGCAACGCGAGGATGCTACAGGTGATTTGCA 2220  
 720 ValThrTyrGluGlyThrIleGluGluAsnGlyGlnGlyMetLeuGlnValAspPheAla 739  
 2221 AATCGTTTGTGGAGGTGGTGTGATCGTGGGAGCTTGTACAGAGAAATCAGATT 2280  
 740 AsnArgPheValGlyGlyValThrSerAlaGlyLeuValGlnGluGluIleArgPhe 759  
 2281 TTAATCAATCTGAAATTCATTTTACGCTGTTCACTGAGGTGCTGATCACAATCAG 2340  
 760 LeuIleAsnProGluLeuIleSerArgLeuPheThrGluValLeuAspHisGlnGlu 779  
 2341 TGTCTTATTATCAGAGTACGAAACAGTACAGTGAATACAGAGTATGCTGAAACTTAT 2400  
 780 CysLeuIleIleThrGlyThrGluGlnTyrSerGluTyrThrGlyTyrAlaGluThrTyr 799  
 2401 CTTGGCCCGAGCATCAAGATGGAGTGAAGGAGGATTCGACGGCGCTGCAGC 2460  
 800 ArgTrpSerArgSerHisGluAspGlySerGluArgAspAspCysGluArgArgCysThr 819  
 2461 GAGATTCGTTGCAATTCAGCACTTCACTTCAGACGCTACTCGATCAGTTTGTGCTGAG 2520  
 820 GluIleValAlaIleAspAlaLeuHisPheArgArgTyrLeuAspGlnPheValProGlu 839  
 2521 AAAGTGACAGCTGAGCTTAAACAGGCTTACTCGGATTCCTCGCTCGAGATTCCTCT 2580  
 840 LysMetArgArgGluLeuAsnLysAlaTyrCysGlyPheLeuArgProGlyValSerSer 859

2581 GAAAATCTTTCTCAGTGGCCAGGAACTGGGGCTGTGGTCTTGGGGGTGACGCT 2640  
 860 GluAsnLeuSerAlaValAlaThrGlyAsnTrpGlyCysGlyAlaPheGlyGlyAspAla 879  
 2641 AGATTAAAGCCCTTAATACAGATCCCTGGCAGCTGCTGGCTGACAGTGGAGTGTAT 2700  
 880 ArgLeuLysAlaLeuIleGlnIleLeuAlaAlaAlaGluArgAspValValTyr 899  
 2701 TTCACTTTGGGACTCAGAGTTGATGAGAGACATTTACAGCATGCACACTTCTCTTACC 2760  
 900 PheThrPheGlyAspSerGluLeuMetArgAspIleTyrSerMetHisIlePheLeuThr 919  
 2761 GAGAGGAAGCTGATGTTGAAAAGTGTACAAGTTATTGCTTAGATCTACAATGAAGAA 2820  
 920 GluArgLysLeuThrValGlyAspValTyrLysLeuLeuLeuArgTyrTyrAsnGluGlu 939  
 2821 TGCAGAAACTGTTCCACCCCTGGACACAGATCAGCTTTATCCATTCATATACCATGCT 2880  
 940 CysArgAsnCysSerThrProGlyProAspIleLysLeuTyrProPheIleTyrHisAla 959  
 2881 GTTGAGTCAAGTGCAGAGACCACTGCATGCCAGACAGAAAGGAGGAGCACC 2931  
 960 ValGluSerCysAlaGluThrAlaAspHisSerGlyGlnArgThrGlyThr 976  
 RESULT 7  
 AAU76012  
 ID AAU76012 standard; protein; 976 AA.  
 XX AC AAU76012;  
 XX 08-MAY-2002 (first entry)  
 DT Human poly(ADP-ribose) glycohydrolase (PARG) protein sequence.  
 DE Human; poly(ADP-ribose) glycohydrolase; PARG; PARG modulator; ADP-ribose;  
 XX adenosine diphosphate-ribose; DNA repair; apoptosis; neoplasia; cancer;  
 KW inherited genetic disease; myocardial infarction; vascular stroke; aging;  
 KW neurodegeneration; Huntington's disease; Parkinson's disease;  
 KW Alzheimer's disease; neurotoxicity.  
 XX Homo sapiens.  
 OS US6333148-B1.  
 PN 25-DEC-2001.  
 PD 30-APR-1999; 99US-00302812.  
 PF 01-MAY-1998; 98US-0083768P.  
 PR (KENT) UNIV KENTUCKY RES FOUND.  
 XX Jacobson MK, Jacobson EL, Ame J, Lin W;  
 FI WPI; 2002-153820/20.  
 DR N-PSDB; ASK14494.  
 XX Screening compounds for modulation of poly(ADP-ribose) glycohydrolase,  
 PT useful potentially for treating diseases associated with DNA damage, e.g.  
 PT cancer.  
 XX Claim 3; Col 55-60; 80pp; English.  
 PS The present invention relates to a new method for screening compounds for  
 XX ability to modulate activity of an enzyme that hydrolyses ADP (adenosine  
 CC inosinate) diphosphate-ribose from an ADP-ribose polymer. The compounds are  
 CC inhibitors or activators of PARG (poly(ADP-ribose) glycohydrolase) and  
 CC are used to treat or prevent any condition associated with DNA damage  
 CC e.g. neoplasia, inherited genetic diseases, myocardial infarction,  
 CC vascular stroke, aging and neurodegeneration e.g. Huntington's,  
 CC Parkinson's or Alzheimer's diseases, or neurotoxicity generally.  
 CC Compounds identified by the new method are more effective than known  
 CC inhibitors and have fewer side effects. The present amino acid sequence

CC represents the human PARG protein of the invention. This protein is one of several PARG proteins (AAU75799 and AAU76012-AAU76015) of the CC invention

XX  
SQ Sequence 976 AA;

## Alignment Scores:

Pred. No.: 0 Length: 976  
Score: 4395.00 Matches: 837  
Percent Similarity: 90.99% Conservative: 52  
Best Local Similarity: 85.67% Mismatches: 78  
Query Match: 64.89% Indels: 10  
DB: 5 Gaps: 6

US-09-302-812-5 (1-3814) x AAU76012 (1-976)

QY 28 ATGAGTGGGGCCCGCTGGAGCCCTGCAGAAAGCG---CGTGGGGCGCGCTGGA 84  
DB 1 MetAsnAlaGlyProGlyCysGluProCysThrLysAlaThrArgTrpGlyAlaAla--- 19  
QY 85 ACTTCTGCGCGACTCCCTCGGACTCCCGGAGCTTCCCTGGCAGGCGAGGGGTCTTC 144  
DB 20 ThrThrSerProAlaAlaSerAspAlaArgSerPheProSerArgGlnArgValLeu 39  
QY 145 GACCCCAAGGAGCTCCCGTCCAGTTTCAGGGTCCCTCCGCTCGCCAGCCTCGCTCG 204  
DB 40 AspProLysAspAlaHisValGlnPheArgValProProSerSerProAlaCysValPro 59  
QY 205 GGGCGGGCGGACCGCACAGAGCAACGCCACTCGTCTGTTTGTTCATAAAGACTATT 264  
DB 60 GlyGlnAlaGlyGlnHisArgGlySerAlaThrSerLeuValPheLysGlnLysThrIle 79  
QY 265 ACTACTGTGATGATTAAGACCAACAGCAGCTGAATCAGAA-----ACTAAA 315  
DB 80 ThrSerTrpMetAspThrLysGlyIleLysThrAlaGluSerGluSerLeuAspSerLys 99  
QY 316 GAAACCAACAATACAGAATTGACTCCATGATGAGTTCTGTGCAGAAAGATACTTTTAC 375  
DB 100 GluAsnAsnThrArgIleGluSerMetSerValGlnLysAspAsnPheTyr 119  
QY 376 CCACATAAGGTGAAATAATGTTCCCTAGCTAAATCTTGATTAATCACCACCA 435  
DB 120 GlnHisAsnValGluLysLeuValAsnValSerGlnLeuSerLeuAspLysSerLeuThr 139  
QY 436 GAAAGAGTTCACAGTATTGTAACCAACAGCAGCTGCGAGTGTGTGCAAGTGGCAGAAT 495  
DB 140 GluLysSerThrGlnLysLeuAsnGlnHisGlnThrAlaAlaMetCysLysTrpGlnAsn 159  
QY 496 GAAGGAAGCATGCAGACAGCTTTTGGCAAGTGAAGCTCCCGCGGGGACTCCGCTACCA 555  
DB 160 GluGlyLysHisThrGluGlnLeuLeuGluSerGluProGlnThrValThrLeuValPro 179  
QY 556 AAGCAGCTTAGTAATGCTTAACATTGTGTAGTACCCACACCTGATCACCAGTGCACCA 615  
DB 180 GluGlnPheSerAsnAlaAsnIleAspArgSerProGlnAsnAspHisSerAspThr 199  
QY 616 GATCATGAAGACAGACAGCAATCAGCAGTTTCTTACACCTATAAACTTGCAATACA 675  
DB 200 AspSerGluGlnAsnArgAspAsnGlnPheLeuThrThrValLysLeuAlaAsnAla 219  
QY 676 AAGCCCAACAGTAGAGATGGCGAGGCGAGA-----AGCAACTGTAAAGTGCAGTGA 726  
DB 220 LysGlnThrThrGluAspGluHisAlaArgGluAlaLysSerHisGlnLysCysSerLys 239  
QY 727 TCTCGGCGAGTCTGTGAAGACTGTACAGGCTGTCAACAGGAGGAGGTGATGTGTACCA 786  
DB 240 SerCysHisProGlyGluAspCysAlaSerCysGlnGlnAspGluIleAspValValPro 259  
QY 787 GAGAGTCTTTGTGCAGATGTGTGCGGAGGACATTTGGAACTGGACCAAAAATGACAAC 846  
DB 260 LysSerProLeuSerAspValGlySerGluAspValGlyThrGlySerLysAsnAspAsn 279  
847 AAATTGACTGGACAGAAAGCAGCCTAGTGTATTCGCCTCCATTTCAGAGAAAGAGTGAG 906

DB 280 LysLeuIleArgGlnGluSerCysLeuGlyAsnSerProProPheGluLysGluSerGlu 299  
QY 907 CCTGAGTCCACCAATGGATGTACCAACTGCAGAAACAGTTGTCAAGATTTCAGAAAGCAGAT 966  
DB 300 ProGluSerProMetAspValAspAsnSerLysAsnSerCysGlnAspSerGluAlaAsp 319  
QY 967 GAAGAAACAAAGTCCAGTCTTTGATGAGCAAGATGAT---CGTTCTCCCAACAGCAAAAT 1023  
DB 320 GluGluThrSerProGlyPheAspGluGlnGluAspGlySerSerSerGlnThrAlaAsn 339  
QY 1024 AAATTTCAAGTTGCCAAGCAAGAGCTGTAGTGGGATCTTAGGAAACAGTATTTCAGT 1083  
DB 340 LysProSerArgPheGlnAlaArgAspAlaAspIleGluPheArgLysArgTyrSerThr 359  
QY 1084 AAGGNAAGTGAAGTTAGATTGCAATTCCTCAATTTGAA---GGAGAAATATATGCTGGACC 1140  
DB 360 LysGlyGlyGluValArgLeuHisPheGlnPheGluGlyGluSerArgThrGlyMet 379  
QY 1141 AGTGACTTAAATGCCAAGCCATCTCTGAAACTCTTCTAGCCTTAAATGTAGAGTGTAGAGT 1200  
DB 380 AsnAspLeuAsnAlaLysLeuProGlyAsnIleSerSerLeuAsnValGluCysArgAsn 399  
QY 1201 TCCAGCAGCATGGAAAAAGGATTTCTAAATACAGATCATTTATGAGATTTTCCAAG 1260  
DB 400 SerLysGlnHisGlyLysLysAspSerLysIleThrAspHisLeuMetArgLeuProLys 419  
QY 1261 TCAGAGGACAGAGAAAGAAACAATGTGAAGTCAGATCAGATCAAGACACAGAAAGAGATT 1320  
DB 420 AlaGluAspArgArgLysGluGlnTrpGluThrLysHisGlnArgThrGluArgLysIle 439  
QY 1321 CCAAAATACATCCCACTTAACCTCCCTCCAGAGAGAAAGTGGCTGGAACTCTCTATTAG 1380  
DB 440 ProLysTyrValProProHisLeuSerProAspLysLysTrpLeuGlyThrProLysGlu 459  
QY 1381 GAAATGAGAAAAATGCTCGTGTGGATCCATTTGCTTCCCTTAAGACCATCTGCAAGT 1440  
DB 460 GluMetArgArgMetProArgCysGlyIleArgLeuProLeuLeuArgProSerAlaAsn 479  
QY 1441 CACAGAGTCACTTCGCGGTAGACTCTCTGAGCAGGAGAGAGTCCGAAACCTTTTCCA 1500  
DB 480 HisThrValThrIleArgValAspLeuLeuArgIleGluValProLysProPhePro 499  
QY 1501 ACACATTACAAAGATTGTGGATAACAAACATGTGAAATGCTTTGTCGGAACAAAC 1560  
DB 500 ThrHisTyrLysAspLeuTrpAspAsnLysHisValLysMetProCysSerGluGlnAsn 519  
QY 1561 TGTGTACCTGTGGAAGATGAGAAATGTGTAGCCGAACCTGCGAGGAGTAGTGGAGCTCAT 1620  
DB 520 LeuTyrProValGluAspGluAsnGlyGluArgThrAlaGlySerArgTrpGluLeuIle 539  
QY 1621 CAGACTGCACCTTCTCAACAAATTCACACACCCAGAACTCGAAGGATCGATTCTGAAA 1680  
DB 540 GlnThrAlaLeuLeuAsnLysPheThrArgProGlnAsnLeuLysAspAlaIleLeuLys 559  
QY 1681 TACAATGTGGCATATTCTTAAGAAATGGGACTTTTACAGCTTTTGGTATTCTGGGATAAG 1740  
DB 560 TyrAsnValAlaTyrSerLysLysTrpAspPheThrAlaLeuIleAspPheTrpAspLys 579  
QY 1741 GTACTTGAAGACGAGAGCCCAACATTATATCATGTCATTTTACCTACATGTTGAAA 1800  
DB 580 ValLeuGluGluAlaGluAlaGlnHisLeuTyrGlnSerIleLeuProAspMetValLys 599  
QY 1801 ATTGCATCTGTCTCCAAATATTTCACCCAGCAATACCACTCTCTGAAACAGAGATG 1860  
DB 600 IleAlaLeuCysLeuProAsnIleCysThrGlnProIleProLeuLeuLysGlnLysMet 619  
QY 1861 AATCATCTGTCCAGATGTTCACAGAAACAGATCGCAGTCTTTTACGTAATGCTTCTTC 1920  
DB 620 AsnHisSerIleThrMetSerGlnGluGlnIleAlaSerLeuLeuAlaAsnAlaPhePhe 639  
QY 1921 TCCACATTTCCCGACGGAATGCCAAGATGAAATCGGAGTATTCCTAGTTACCCAGCAT 1980

Db 640 CysThrPheProArgAsnAlaLysMetLysSerGluTyrSerSerTyrProAspIle 659  
 Qy 1981 AACTTCATCGTGTGTTGAAGGAGCTTCATCAGGAACCCAGAAACTGAAACACTC 2040  
 Db 660 AsnPheAsnArgLeuPheGluGlyArgSerArgLysProGluLysLeuLysThrLeu 679  
 Qy 2041 TTCTGCTACTTTCGAAGAGTCACAGAGAAACCTACAGATTGGTGACATTTACAGA 2100  
 Db 680 PheCysTyrPheArgArgValThrGluLysLysProThrGlyLeuValThrPheThrArg 699  
 Qy 2101 CAGAGCTTTGAAGATTTCCAGAAATGGGAAGGTGTGAAAGCTCTTGACACGCTTACAC 2160  
 Db 700 GlnSerLeuGluAspPheProGluTyrGluArgCysGluLysProLeuThrArgLeuHis 719  
 Qy 2161 GTCACTTACAGGGTACATAGAGGCAACCGCGAGGCATGCTACAGGTGGATTGCA 2220  
 Db 720 ValThrTyrGluGlyThrIleGluGluAsnGlyGlnGlyMetLeuGlnValAspPheAla 739  
 Qy 2221 AATCGTTTGTGGAGTGTGTGACTGCTGCGGACTTGTACAGAGAAATCAGATTT 2280  
 Db 740 AsnArgPheValGlyGlyValThrSerAlaGlyLeuValGlnGluIleArgPhe 759  
 Qy 2281 TTAATCAATCCTGAATGATTTTCACGGCTGTTCACGTGAGTGTGATGATCAATGAG 2340  
 Db 760 LeuIleAsnProGluLeuIleIleSerArgLeuPheThrGluValLeuAspHisAsnGlu 779  
 Qy 2341 TGTCTTATTATCAGGTTACTGAACAGTACAGTACAGTACAGGCTATGCTGAATCTAT 2400  
 Db 780 CysLeuIleIleThrGlyThrGluGlnTyrSerGluTyrThrGlyTyrAlaGluThrTyr 799  
 Qy 2401 CGTTGGGCCCAAGCCATGAGATGGGAGTGGAGTGGAGACGATTCGACGCGCTGCACG 2460  
 Db 800 ArgTrpSerArgSerHisGluAspGlySerGluArgAspAspCysGluArgCysThr 819  
 Qy 2461 GAGATCGTTCATGAGCGCTTCATCTCAGAGCGTACCTGATCGATCGATTTGCTGCTGAG 2520  
 Db 820 GluIleValAlaIleAspAlaLeuHisPheArgArgTyrLeuAspGlnPheValProGlu 839  
 Qy 2521 AAGTGAGAGCTGAGCTTAAACAGCTTACTCGGATTCCTCGCTGCGAGTTCCTTCT 2580  
 Db 840 LysMetArgArgGluLeuAsnLysAlaTyrCysGlyPheLeuArgProGlyValSerSer 859  
 Qy 2581 GAAATCTTTCGACGTGCGCACCGGAACTCGGGCTGTGCTGCTTGGGGGTGAGCT 2640  
 Db 860 GluAsnLeuSerAlaValAlaThrGlyAsnTrpGlyCysGlyAlaPheGlyGlyAspAla 879  
 Qy 2641 AGATTAAAGCTTAATACAGATCTGCGAGCTGCTGCGGTGAAACGTCGAGTGTGTTAT 2700  
 Db 880 ArgLeuLysAlaLeuIleGlnIleLeuAlaAlaAlaGluArgAspValValTyr 899  
 Qy 2701 TTCACCTTGGGACTCAGAGTGTGATGAGACATTTTACAGTCGACACTTCTTCTTACC 2760  
 Db 900 PheThrPheGlyAspSerGluLeuMetArgAspIleTyrSerMetHisIlePheLeuThr 919  
 Qy 2761 CAGAGGAGCTGGATGTTGGAAAAGTGTACAGTTTATGCTTAGATACATACATGAAGA 2820  
 Db 920 GluArgLysLeuThrValGlyAspValTyrLysLeuLeuLeuArgTyrTyrAsnGluGlu 939  
 Qy 2821 TGCAGAAACTGTTCCACCCCTGGACGACATCAAGCTTTTATCCATTCATATACATGCT 2880  
 Db 940 CysArgAsnCysSerThrProGlyProAspIleLysLeuTyrProPheIleTyrHisAla 959  
 Qy 2881 GTTGGAGTCAAGTGCAGAGACCACTACATGCGCAGGACAGAGGCGAGCC 2931  
 Db 960 ValGluSerCysAlaGluThrAlaAspHisSerGlyGlnArgThrGlyThr 976

RESULT 8

ASG72279  
 ID ABG72279 standard; protein; 976 AA.

XX ABG72279;

XX DT 13-MAR-2003 (first entry)

XX DE  
 XX KW Human poly (ADP-ribose) glycohydrolase (PARG) enzyme.  
 KW Human; poly (ADP-ribose) glycohydrolase; PARG; PARG expression;  
 KW cellular response; DNA damage; neoplastic disorder; inducing agent;  
 KW oxidative stress; neoplastic disorder; myocardial infarction;  
 KW vascular stroke; neurodegenerative disease; Alzheimer's disease;  
 KW Parkinson's disease; Huntington's disease; inborn genetic error;  
 KW reperfusion; ischaemia; aging; neurotoxicity; PARG activity; cytostatic;  
 KW neuroprotective; nontoxic; antiparkinsonian; cardiant; vasotropic;  
 KW anticonvulsant; cerebroprotective; enzyme.  
 XX OS Homo sapiens.  
 XX US US2002132328-A1.  
 XX PD 19-SEP-2002.  
 XX PF 09-OCT-2001; 2001US-00973451.  
 XX PR 01-MAY-1998; 98US-0083768P.  
 XX PR 30-APR-1999; 99US-00302812.  
 XX PA (JACO/) JACOBSON M K.  
 PA (JACO/) JACOBSON E L.  
 PA (AMEJ/) AME J.  
 PA (LINW/) LIN W.  
 XX PI Jacobson MK, Jacobson EL, Ame J, Lin W;  
 XX WI WPI; 2003-155895/15.  
 XX DR N-PSDB; ABX14478.  
 XX PT New nucleic acid molecule encoding a polypeptide with poly (ADP-ribose)  
 PT glycohydrolase activity, for preventing, treating, or ameliorating a  
 PT disease condition, e.g. neoplastic disorder, myocardial infarction or  
 PT vascular stroke.  
 XX PS Claim 28; Fig 16; 86pp; English.  
 CC The present invention relates to the isolation of poly (ADP-ribose)  
 CC glycohydrolase (PARG) from several species, and the polynucleotide  
 CC sequences encoding them. Methods for inhibiting PARG expression or  
 CC overexpressing PARG are also disclosed. PARG is involved in the cellular  
 CC response to DNA damage, and is associated with the body's response to  
 CC neoplastic disorder inducing agents and oxidative stress. The  
 CC polynucleotide sequences encoding PARG and PARG modulators are useful for  
 CC preventing, treating, or ameliorating diseases such as neoplastic  
 CC disorders, myocardial infarction, vascular stroke, neurodegenerative  
 CC disorders (e.g. Alzheimer's disease, Parkinson's disease or Huntington's  
 CC disease), inborn genetic errors, reperfusion following ischaemia, aging,  
 CC and neurotoxicity. The polynucleotide sequences are also useful in gene  
 CC therapy. The methods are useful for identifying an agent that modulates  
 CC PARG activity, identifying a mutant PARG allele in an individual, or  
 CC screening candidate molecules for PARG modulating activity. The present  
 CC sequence represents human PARG enzyme  
 XX SQ Sequence 976 AA;  
 Alignment Scores:  
 Pred. No.: 0 Length: 976  
 Score: 4395.00 Matches: 837  
 Percent Similarity: 90.99% Conservative: 52  
 Best Local Similarity: 85.67% Mismatches: 78  
 Query Match: 64.89% Indels: 10  
 DB: 6 Gaps: 6  
 US-09-302-812-5 (1-3814) x ABG72279 (1-976)  
 Qy 28 ATGAGTGGGCGCCCGCTGGAGCCCTGCACGAAGCG---CGCTGGGCGCGCTGGA 84  
 Db 1 MetAsnAlaGlyProGlyCysGluProCysThrLysAlaThrArgTrpGlyAlaAla--- 19

85 ACTTCTGCGGACTGCTCCGAGCTCCCGAGCTTCCCTGGCAGGAGGCGGTGTTCTC 144  
120 ThrThrSerProAlaAlaSerAspAlaArgSerPheProSerArgGlnArgValLeu 39  
145 GACCCCAAGAGCGCTCCCGTCCAGTTCAGGGTCCCTCCGTCGCGCAGCTCGCTCG 204  
40 AspProLysAspAlaHisValGlnPheArgValProProSerSerProAlaCysValPro 59  
205 GGGGGGGGGCCGACAGAGGCAAGCCACCTCGTTGTTTTCACAAAGAGACTATT 264  
60 GlyGlnAlaGlyGlnHisArgGlySerAlaThrSerLeuValPheLysGlnLysThrIle 79  
265 ACTACTTGATGATGATACTAAAGGACCAACAGACAGCTGAATCAGAA-----AGTAAA 315  
80 ThrSerTrpMetAspThrLysGlyIleLysThrAlaGluSerLysLeuAspSerLys 99  
316 GAAACAAACAATACAGAATTGATCCATGATGAGTTCGTGTCAGAAAGATACATTTTAC 375  
100 GluAsnAsnThrArgIleGluSerMetMetSerSerValGlnLysAspAsnPheTyr 119  
376 CCACATAAGGTGGAAAATGGAAAATGTTCTCAGCTAAATCTTGATAAATCACCCACA 435  
120 GlnHisAsnValGluLysLeuValAsnValSerGlnLeuSerLeuAspLysSerLeuThr 139  
436 GAAAAGAGTTCAGATATTGTAACCAACAGACACTGCGAGTGTGTGCAAGTGGCAGAAT 495  
140 GluLysSerThrGlnTyrLeuAsnGlnHisGlnThrAlaAlaMetCysLysTrpGlnAsn 159  
496 GAAGGAAGCATGCGAAGCAGCTTTTGGCAAGTGAAGTCCCGCGGGGAGCTCGCTACCA 555  
160 GluGlyLysHisThrGluGlnLeuLeuGluSerGluProGlnThrValThrLeuValPro 179  
556 AAGCAGCTTAGTAAGTAACTAAGTGTGTCAGTCACTCCACACACATGATGACACACAGTACACA 615  
180 GluGlnPheSerAsnAlaAsnIleAspArgSerProGlnAsnAspAspHisSerAspThr 199  
616 GATCATGAAGAAAGACAGACAAATCAGCAGTGTCTTACACCTATATAAACTGCGAAATACA 675  
200 AspSerGluGluAsnArgAspGlnGlnPheLeuThrValLysLeuAlaAsnAla 219  
676 AAGCCAACAGTAGGAGATGGCGAGGCGACA-----AGCAACTGTAAAGTCCAGTGA 726  
220 LysGlnThrThrGluAspGluHisAlaArgGluAlaLysSerHisGlnLysCysSerLys 239  
727 TCTCCCGAGTGTGAAAGACTGTACAGGCTGTCAACAGAGAGGTGATGCTACCA 786  
240 SerCysHisProGlyGluAspCysAlaSerCysGlnGlnAspGluIleAspValValPro 259  
787 GAGAGTCTTTGTGATGTTGTGCGGAGGACATTTGGAACCTGGACCAAAATGACAAAC 846  
260 LysSerProLeuSerAspValGlySerGluAspValGlyThrGlySerLysAsnAspAsn 279  
847 AAATTGACTGGCAAGAAAGACGCTTAGTGATTCGCTCCATTTGAGAAAGAAAGTAG 906  
280 LysLeuIleArgGlnGluSerCysLeuGlyAsnSerProPheGluLysGluSerGlu 299  
907 CTGAGTCCCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 966  
300 ProGluSerProMetAspValAspAsnSerLysAsnSerCysGlnAspSerGluAlaAsp 319  
967 GAAGAAACAAAGTCCAGTCTTTGATGAGCAAGATGAT---CGTTCTCTCCCAAAACAGCAAAAT 1023  
320 GluGluThrSerProGlyPheAspGluGlnGluAspGlySerSerSerGlnThrAlaAsn 339  
1024 AAATTTCAAGTTGCCAAGCAGAGAGCTGATGCGGATCTTAGGAAACGGTATTGACT 1083  
340 LysProSerArgPheGlnAlaArgAspAlaAspIleGluPheArgLysArgTyrSerThr 359  
1084 AAGGGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 1140  
360 LysGlyGlyGluValArgLeuHisPheGlnPheGluGlyGlyGluSerArgThrGlyMet 379  
1141 AGTGACTTAAATGCCAAGCATCTCGAAACTCTTCTAGCTTAAATGAGGTGAGAGT 1200

380 AsnAspLeuAsnAlaLysLeuProGlyAsnIleSerSerLeuAsnValGluCysArgAsn 399  
1201 TCACAGCAGCATGGAAAAAGGATTTCTAAAAATACAGATCATTTTCATGAGAAATTCAG 1260  
400 SerLysGlnHisGlyLysLysAspSerLysIleThrAspHisLeuMetArgLeuProLys 419  
1261 TCAGAGGACAGAGAAAGAAACAATCTGAAGTCAGACATCAAGACACAGAAAGAGAGATT 1320  
420 AlaGluAspArgArgLysGluGlnTrpGluThrLysHisGlnArgThrGluArgLysIle 439  
1321 CCAAAATACATCCCACTAAACCTCCCTCCAGAGAGAAAGTGGCTGGGAATCTCTATTGAG 1380  
440 ProLysTyrValProProHisLeuSerProAspLysLysTrpLeuGlyThrProLysGlu 459  
1381 GAAATGAGAAATATGCTCGGTGGGATCCATTTGCTTCCCTTAAAGACAATCTGCAAGT 1440  
460 GluMetArgArgMetProArgCysGlyIleArgLeuProLeuLeuArgProSerAlaAsn 479  
1441 CACACAGTGAAGTTCGGGTAGACCTTCTGAGCAGGAGAGAGTTCGGAACCTTTTCCA 1500  
480 HisThrValThrIleArgValAspLeuLeuArgAlaGlyGluValProLysProPhePro 499  
1501 ACACATTACAAAGATTGTGGGATACAAACATGTGAAATGCTTTGTTCCGGAACAAAC 1560  
500 ThrHisTyrLysAspLeuTrpAspAsnLysHisValLysMetProCysSerGluGlnAsn 519  
1561 TTGTACCTCTGCGAAGATGAGAAATGGTGGAGCGAATCTGAGGAGTAGGTGGAGCTCAT 1620  
520 LeuTyrProValGluAspGluAsnGlyGluArgThrAlaGlySerArgTrpGluLeuIle 539  
1621 CAGAGTGCATCTCAACAAATTCACAGACCCCAAGACTTGAAGGATCGATTCGAGAA 1680  
540 GlnThrAlaLeuLeuAsnLysPheThrArgProGlnAsnLeuLysAspAlaIleLeuLys 559  
1681 TACAATTTGGCATATTCTAAGAAATGGGACTTTTACAGCTTTTGGTATTCTGGGATAAG 1740  
560 TyrAsnValAlaTyrSerLysLysTrpAspPheThrAlaLeuIleAspPheTrpAspLys 579  
1741 GTACTTGAAGAGCAGAGCCCAACATTTATATCATGTCATTTTACCTGACATGTTGAA 1800  
580 ValLeuGluGluAlaGluAlaGlnHisLeuTyrGlnSerIleLeuProAspMetValLys 599  
1801 ATTGCATCTGCTCCCAATATTTGCCACCCAGCAATACCACTCTCTGAAACAGAGATG 1860  
600 IleAlaLeuCysLeuProAsnIleCysThrGlnProIleProLeuLeuLysGlnLysMet 619  
1861 AATCATTTCTGTCAGATGTCACAGAAACAGATCGCCAGTCTTTTACGTAATGCTTTCTTC 1920  
620 AsnHisSerIleThrMetSerGlnGluGlnIleAlaSerLeuLeuAlaAsnAlaPhePhe 639  
1921 TCACATTTCCCGCAGGATGCCAGATGAATCGAGTATTTCTAGTTACCCAGACATT 1980  
640 CysThrPheProArgArgAsnAlaLysMetLysSerGluTyrSerSerTyrProAspIle 659  
1981 RAATTCAATCGGTCTTTGAAGACGTTTCATCAAGAAACACAGAAACACGAAACACATC 2040  
660 AsnPheAsnArgLeuPheGluArgSerArgLysProGluLysLeuLysThrLeu 679  
2041 TTCTCTACTTTTCGAGAGTCACAGAAAAAACCTACAGGATTCGTCATTTTCAAGA 2100  
680 PheCysTyrPheArgArgValThrGluLysLysProThrGlyLeuValThrPheThrArg 699  
2101 CAGACTCTGAGATTTTCCGAATGGGAAGTCTGAAAGCTCTGACAGCTTACAC 2160  
700 GlnSerLeuGluAspPheProGluTrpGluArgCysGluLysProLeuThrArgLeuHis 719  
2161 GTCACTTACAGGGTACCATAGAGCAACCGCCCGAGGATCTCTACAGGTGATTTTGA 2220  
720 ValThrTyrGluGlyThrIleGluGluAsnGlyGlnGlyMetLeuGlnValAspPheAla 739  
2221 AATCCTTTTGTGGAGGTGGTGTGACTGGTGGGACTTGTACAAAGAAATTCAGATT 2280





Db	881	ArgLeuLysAlaLeuIleGlnIleLeuAlaAalaValaGluArgAspValValTyr	900
QY	2701	TTCACTTTGGGACTCAGAGTGTGATCAGAGACATTTACAGCATGCACACTTTCTTACC	2760
Db	901	PheThrPheGlyAspSerGluLeuMetArgAspIleTyrSerMetHisThrPheLeuThr	920
QY	2761	GAGAGGAAGCTGGATGTTGGAAAAGTGTACAAGTTATTGCTTTAGATCTACTCAATGAAGAA	2820
Db	921	GluArgLysLeuThrValGlyGluValTyrLysLeuLeuLeuArgTyrTyrAsnGluGlu	940
QY	2821	TGCAGAACTGTTCCACCCCTGGACACAGACATCAAGCTTTTATCCATTCATATACCATGCT	2880
Db	941	CysArgAsnCysSerThrProGlyProAspIleLysLeuTyrProPheIleTyrHisAla	960
QY	2881	GTTGAGTCAAGTGCAGAGACCACTGCATGCACGACAGAAAGGCAGGC	2928
Db	961	ValGluSerCysThrGlnThrThrAsnGlnProGlyGlnArgThrGly	976
RESULT 10			
AAU76020			
ID	AAU76020 standard; protein; 977 AA.		
XX			
AC	AAU76020;		
XX			
DT			
XX			
XX			
DE	08-MAY-2002 (first entry)		
XX			
XX	Bovine poly (ADP-ribose) glycohydrolase (PARG) protein sequence.		
KW	Cow; poly (ADP-ribose) glycohydrolase; PARG; DNA repair; ADP-ribose;		
KW	adenosine diphosphate-ribose; neoplastic disorder; adenocarcinoma;		
KW	leukaemia; lymphoma; melanoma; myeloma; sarcoma; teratocarcinoma;		
KW	hyperplasia; hypertrophy; reperfusion; ischaemia; heart attack; stroke;		
KW	neurodegenerative disease; neurological disorder; Alzheimer's disease;		
KW	Huntington's disease; Parkinson's disease.		
OS	Bos taurus.		
XX			
Key	Location/Qualifiers		
FT	Region	601..617	
FT		/note= "Represents PARG oligopeptide #1"	
FT	Region	761..770	
FT		/note= "Represents PARG oligopeptide #2"	
FT	Region	771..801	
FT		/note= "Represents PARG oligopeptide #3"	
FT	Region	849..880	
FT		/note= "Represents PARG oligopeptide #4"	
XX			
PN	USG337202-B1.		
XX			
PD	08-JAN-2002.		
XX			
PF	23-FEB-2000; 2000US-00511477.		
XX			
PR	01-MAY-1998; 98US-0083768P.		
PR	30-APR-1999; 99US-00302812.		
XX			
PA	(KENT ) UNIV KENTUCKY RES FOUND.		
XX			
PI	Jacobson MK, Jacobson EL, Ame J, Lin W;		
XX			
DR	WPI; 2002-163240/21.		
DR	N-PSDB; ABK14931.		
XX			
PT	Novel isolated and purified poly (ADP-ribose) glycohydrolase protein		
PT	which catalyses release of ADP-ribose from ADP ribose polymer, useful for		
PT	treating neoplastic and neurological disorders, heart attack and stroke.		
XX			
PS	Claim 2; Col 47-52; 81pp; English.		
XX			
CC	The present invention relates to a new poly (ADP-ribose) glycohydrolase		
CC	(PARG) protein which catalyses release of ADP-ribose from an ADP		
CC	(adenosine diphosphate)-ribose polymer. The PARG molecule of the		
CC	invention is useful for generating antibodies and can be inhibited or		



241 AlaCysHisProAlaGluAlaCysAlaGlyCysGlnGlnGluGluThrAspValValSer 260  
 787 GAGAGTCCTTTGTGTCAGATGTTGGTCCGAGGACATTTGGAATCGGACCAAAATGACAAAC 846  
 261 GluSerProLeuSerAspThrGlySerGluAspValGlyThrGlyLeuLysAsnAlaAsn 280  
 847 AAATGACTGGCAAGAACAGCAGCTAGGTGATTCCTCCATTTGAGAAAGAAAGTGAAG 906  
 281 ArgLeuAsnArgGlnGluSerLeuGlyAsnSerProPheGluLysGluSerGlu 300  
 907 CTGAGTCACCAATGATGTAGCAACTCGAGAAACAGTTGTCAAGATTCAAGAGCAGAT 966  
 301 ProGluSerProMetAspValAspAsnSerLysAsnSerCysGlnAspSerGluAlaAsp 320  
 967 GAAGAAACAAAGTCCAGTCTTGTATGAGCAAGATGATCGTTCCTCC---CAAAACAGCAAAAT 1023  
 321 GluGluThrSerProGlyPheAspGluGlnGluAspSerSerAlaGlnThrAlaAsn 340  
 1024 AAATTTCAAGTTGCCAAGCAAGAGAGCTGATGGCGATCTTAGGAAACGGTATTGACT 1083  
 341 LysProSerArgPheGlnProArgGluAlaAspThrGluLeuArgLysArgSerAla 360  
 1084 AAGGGAAGTGAAGTGAATTCATTTCCAAATTGAA--GGAGAAATAATGCTGGGACC 1140  
 361 LysGlyGlyGluLeuLeuLeuHisPheGlnPheGluGlyGlyGluSerArgAlaGlyMet 380  
 1141 AGTGACTTAAATGCCAAGCCATCTCGAAACTCTTCTAGCCTTAAATGAGAGTGAAGT 1200  
 381 AsnAspValAsnAlaLysArgProGlySerThrSerSerLeuAsnValGluCysArgAsn 400  
 1201 TCACAGCAGCATGGAAAAGGATTCCTAAATTAAGATCATTTCTAGAAATTTCCAAAG 1260  
 401 SerLysGlnHisGlyArgLysAspSerLysIleThrAspHisPheMetArgValProLys 420  
 1261 TCAGAGGACAGAGAAAGAAACAATCTGAAGTCAGACATCAAGACAGAAAGAGATG 1320  
 421 AlaGluAspLysArgLysGluGlnCysGluMetLysHisGlnArgThrGluArgLysIle 440  
 1321 CCAAAATACATCCCACTAACCTCCCTCCAGAGAAAGTGGCTGGGAACTCCTATTGAG 1380  
 441 ProLysTyrIleProProHisLeuSerProAspLysLysTrpLeuGlyThrProLysGlu 460  
 1381 GAAATGAGAAAATGCTCGGTGGGATCCATTTCCCTTAAAGACATCTGCAAGT 1440  
 461 GluMetArgArgMetProArgCysGlyIleArgLeuProLeuArgProSerAlaAsn 480  
 1441 CACAGAGTCACTGCTCGGTGAGACCTCTGAGAGCAGGAGAGTTCGGAACCTTTTCCA 1500  
 481 HisThrValThrIleArgValAspLeuLeuLeuArgIleGlyGluValProLysProPhePro 500  
 1501 ACACATTAACAAGATTGTGGGATAACAACATGTGAAAATGCTTTGTTCCGGAACAAAAC 1560  
 501 ThrHisPheLysAspLeuTrpAspAsnLysHisValLysMetProCysSerGluGlnAsn 520  
 1561 TTGTACCTGTGAGAGATGAGATGTGAGCGAACTGCGAGGAGTAGGTGGAGCTCAT 1620  
 521 LeuTyrProValGluAspGluAsnGlyGluArgAlaAlaGlySerArgTrpGluLeuIle 540  
 1621 CAGACTGCACTCTCAACAAATTCACAGCCCACTGAGATCTGAAGATCGGATCTCGAAA 1680  
 541 GlnThrAlaLeuLeuAsnArgLeuThrArgProGlnAsnLeuLysAspAlaIleLeuLys 560  
 1681 TACAATGTGGCATTTCTAAGAAATGGGACTTTTACAGCTTTGTGATTCTCTGGGATAAG 1740  
 561 TyrAsnValAlaTyrSerLysLysTrpAspPheThrAlaLeuIleAspPheTrpAspLys 580  
 1741 GTACTTGAGAGACAGAGCCCAACATTTATATATCATCTCATTTTACCTGACATGGTGAAA 1800  
 581 ValLeuGluGluAlaGluAlaGlnHisLeuTyrGlnSerIleLeuProAspMetValLys 600  
 1801 ATTGCACCTCTGCTGCCAAATATTTCCACCACCAATACCACTCTCTGAAAACAGAGATG 1860

601 IleAlaLeuCysLeuProAsnIleCysThrGlnPnProIleProLeuLeuLysGlnLysMet 620  
 1861 AATCATTTGTCACGATGTCAGAGAACAGATCGGCAGTCTTTTACGTAATGCTTTCTTC 1920  
 621 AsnHisSerIleThrMetSerGlnGluGlnIleAlaSerLeuLeuAlaAsnAlaPhePhe 640  
 1921 TGCACATTTCCCGGACGGAATGCCAGATGAATCGAGTATTCAGTATACCCAGACATT 1980  
 641 CysThrPheProArgAsnAlaLysMetLysSerGluTyrSerSerTyrProAspIle 660  
 1981 AACTTCAATTCGCTGTTTGAAGGACCTTCATCAAGGAAACCCAGAAAACATGAAAAACATC 2040  
 661 AsnPheAsnArgLeuPheGluGlyArgSerSerArgLysProGluLysLeuLysThrLeu 680  
 2041 TTCTGCTACTTTCCGAAGAGTCAAGAGAAAACCTACAGGATTTGGTGTGACATTACAGA 2100  
 681 PheCysTyrPheArgArgValThrGluLysLysProThrGlyLeuValThrPheThrArg 700  
 2101 CAGAGTCTTGAAGATTTCCAGAATGGGAAAGTGTGAAAAGCCTCTGACAGCTTACAC 2160  
 701 GlnSerLeuGluAspPheProGluTrpGluArgCysGluLysLeuLeuThrArgLeuHis 720  
 2161 GTCACCTTACGAGGTACCATAGAGGCAACGGCCGAGGATGCTACAGGTGGATTTCGA 2220  
 721 ValThrTyrGluGlyThrIleGluGlyAsnGlyGlnGlyMetLeuGlnValAspPheAla 740  
 2221 AATCGTTTGTGGAGGTGTGTGACTGGTGGGAGTCTGTACAGAGAAATACAGATT 2280  
 741 AsnArgPheValGlyGlyValThrSerAlaGlyLeuValGlnGluGluIleArgPhe 760  
 2281 TTAATCAATCCTGAATGATGATTTTCAGGCGTGTTCACGTGAGTGTGATCAATGAG 2340  
 761 LeuIleAsnProGluLeuIleValSerArgLeuPheThrGluValLeuAspHisAsnGlu 780  
 2341 TCTCTTATTATCACAGGTACTCAACAGTACAGTGAATACACAGGCTATCTGAAACTTAT 2400  
 781 CysLeuIleIleThrGlyThrGluGlnTyrSerGluTyrThrGlyTyrAlaGluThrTyr 800  
 2401 CTTTGGGCCCCGAAGCATGAAGATGGAGTGAAGAGAGCATGGAGCGGCGCTGCAG 2460  
 801 ArgTrpAlaArgSerHisGluAspArgSerGluArgAspAspTrpGlnArgArgThrThr 820  
 2461 GAGATCGTTCATTCAGCAGCTTCACATTCAGACGCTACCTCGATCAGTTGTGCTGAG 2520  
 821 GluIleValAlaIleAspAlaLeuHisPheArgArgTyrLeuAspGlnPheValProGlu 840  
 2521 AAAGTGAGACGTGAGCTTAAACAAGGCTTACTCGGATTCCTCGCTCGAGTTCCTCT 2580  
 841 LysIleArgArgGluLeuAsnLysAlaTyrCysGlyPheLeuArgProGlyValSerSer 860  
 2581 GAAATCTTTTCTGAGTGCCACGGGAAACTCGGGCTGTGGTGGCTTTGGGGTGAGCT 2640  
 861 GluAsnLeuSerAlaValAlaThrGlyAsnTrpGlyCysGlyAlaPheGlyGlyAspAla 880  
 2641 AGATTAAGACCTTAATACAGATCTCGCAGCTGTGCGGCTGAACGTGACGCTGGTTAT 2700  
 881 ArgLeuLysAlaLeuIleGlnIleLeuAlaAlaValAlaGluArgAspValValTyr 900  
 2701 TTCACCTTTGGGACTCAGAGTTGATGAGACATTTACAGCATGACATTCCTTCTTACC 2760  
 901 PheThrPheGlyAspSerGluLeuMetArgAspIleTyrSerMetHisThrPheLeuThr 920  
 2761 GAGAGAGCTGGATGTTGGAAAAGTGTACAGATTATTCCTAGATACATACATGAGAA 2820  
 921 GluArgLysLeuThrValGlyGluValTyrLysLeuLeuLeuArgTyrTyrAsnGluGlu 940  
 2821 TCGAGAACTGTTCACCCCTCGACAGACATCAAGCTTTATCCATTATATACCATGCT 2880  
 941 CysArgAsnCysSerThrProGlyProAspIleLysLeuTyrProPheIleTyrHisAla 960  
 2881 GTTGAGTCAAGTGCAGACCACTGACATGCCAGGACGAGGACGAGC 2928  
 961 ValGluSerCysThrGlnThrThrAsnGlnProGlyGlnArgThrGly 976



RESULT 11  
AAU75799  
ID AAU75799 standard; protein; 977 AA.  
AC AAU75799;  
XX  
DT 08-MAY-2002 (first entry)  
XX  
DE Bovine poly(ADP-ribose) glycohydrolase (PARG) protein sequence.  
XX  
KW Cow; poly(ADP-ribose) glycohydrolase; PARG; PARG modulator; ADP-ribose;  
KW adenosine diphosphate-ribose; DNA repair; apoptosis; neoplasia; cancer;  
KW inherited genetic disease; myocardial infarction; vascular stroke; aging;  
KW neurodegeneration; Huntington's disease; Parkinson's disease;  
KW Alzheimer's disease; neurotoxicity.  
XX  
OS Bos taurus.  
XX  
PH Key Location/Qualifiers  
FT Region 601..617  
FT /note= "Represents PARG oligopeptide #1"  
FT Region 761..770  
FT /note= "Represents PARG oligopeptide #2"  
FT Region 771..801  
FT /note= "Represents PARG oligopeptide #3"  
FT Region 849..880  
FT /note= "Represents PARG oligopeptide #4"  
XX  
US6333148-B1.  
XX  
25-DEC-2001.  
XX  
30-APR-1999; 99US-00302812.  
XX  
01-MAY-1998; 98US-0083768P.  
XX  
(KENT) UNIV KENTUCKY RES FOUND.  
XX  
Jacobson MK, Jacobson EL, Ame J, Lin W;  
XX  
WPI: 2002-153820/20.  
DR N-PSDB; ABK14493.  
XX  
Screening compounds for modulation of poly(ADP-ribose) glycohydrolase,  
PT useful potentially for treating diseases associated with DNA damage, e.g.  
PT cancer.  
XX  
Claim 3; Col 45-52; 80pp; English.  
XX  
The present invention relates to a new method for screening compounds for  
CC ability to modulate activity of an enzyme that hydrolyses ADP (adenosine  
CC diphosphate)-ribose from an ADP-ribose polymer. The compounds are  
CC inhibitors or activators of PARG (poly(ADP-ribose) glycohydrolase) and  
CC are used to treat or prevent any condition associated with DNA damage  
CC e.g. neoplasia, inherited genetic diseases, myocardial infarction,  
CC vascular stroke, aging and neurodegeneration e.g. Huntington's,  
CC Parkinson's or Alzheimer's diseases, or neurotoxicity generally.  
CC Compounds identified by the new method are more effective than known  
CC inhibitors and have fewer side effects. The present amino acid sequence  
CC represents the bovine PARG protein of the invention. This protein is one  
CC of several PARG proteins (AAU75799 and AAU76012-AAU76015) of the  
XX invention  
SQ Sequence 977 AA;  
Alignment Scores:  
Pred. No.: 0 Length: 977  
Score: 4332.50 Matches: 820  
Percent Similarity: 89.86% Conservative: 57  
Best Local Similarity: 84.02% Mismatches: 90  
Query Match: 63.97% Indels: 9  
DB: 5 Gaps: 5

US-09-302-812-5 (1-3814) x AAU75799 (1-977)  
QY 28 ATGAGTCGGGGCCCGGCTGGAGCCCTGCACGAAA---GGCGCTGGGGCGCGCTGGA 84  
DB 1 MetSerAlaGlyProGlyCysGluProCysThrLysArgProArgTTPAspAlaAla 20  
QY 85 ACTTCTGCGCCGACTGCTCGGACTCCCGGAGTCTCCCTGGCAGGCGAGCGGTCTC 144  
DB 21 ThrSerProAlaAlaSerAspAlaArgSerPheProGlyArgGlnArgValLeu 40  
QY 145 GACCCCAAGACGCTCCCGTCCAGTTCAGGTCCTCCCTCCCTCCGCGAGCTCGTCTCG 204  
DB 41 AspSerLysAspAlaProValGlnPheArgValProProSerSerSerGlyCysAlaLeu 60  
QY 205 GGGCGGGCGGACCGCAGAGGCAAGCCACCTCGTTCGTTTTCATCAACAAAGACTATT 264  
DB 61 GlyArgAlaGlyGlnHisArgGlySerAlaThrSerLeuValPheLysGlnLysThrIle 80  
QY 265 ACTACTTGGATGATTAAGGACCAACAGACAGCTGAATCAGAA-----AGTAAA 315  
DB 81 ThrSerTrpMetAspThrLysGlyIleLysThrValGluSerGluSerLeuHisSerLys 100  
QY 316 GAAACCAACAATACAGAAATTGACTCCATCATGATGTTCTGTGCAAGAAAGATACTTTTAC 375  
DB 101 GluAsnAsnAsnThrArgGluGluSerMetSerSerValGlnLysAspAsnPheTyr 120  
QY 376 CCACATAAGGTGAAATAATGTTCTCCTCAGCTAAATCTTGATAATCAACCCACA 435  
DB 121 GlnHisAsnMetGluLysLeuGluAsnValSerGlnLeuGlyPheAspLysSerProVal 140  
QY 436 GAAAGAGTTTCACAGTATTGTAACCAACACAGACAGCTGGAGTGTGTGCAAGTGCAGAAT 495  
DB 141 GluLysGlyThrGlnTyrLeuLysGlnHisGlnThrAlaAlaMetCysLysTrpGlnAsn 160  
QY 496 GAAGGGAAGCATGCAGAACAGCTTTTGGCAAGTGAGCTCCCGCGGGAGTCCGCTACCA 555  
DB 161 GluGlyProHisSerGluArgLeuGluSerGluProProAlaValThrLeuValPro 180  
QY 556 AAGCAGCTTAGTAACTGCTAACTTGTGCTACCTACCCACACTGATGACCACTGACACA 615  
DB 181 GluGlnPheSerAsnAlaAsnValAspGlnSerSerProLysAspAspHisSerAspThr 200  
QY 616 GATCATGCAAGACACAGACAGCAATCAGCTTTTACACCTATACCAATTCAGAAATACA 675  
DB 201 AsnSerGluGluSerArgAspAsnGlnGlnPheLeuThrHisValLysLeuAlaAsnAla 220  
QY 676 AAGCCAACAGTAGGATGGGAG-----GCCAGAAAGCAACTGTAAGTGCAGTGA 726  
DB 221 LysGlnThrMetGluAspGluGlnGlyArgGluAlaArgSerHisGlnLysCysGlyLys 240  
QY 727 TCTGCGCAGTCTGTGAAAGACTGTACAGGCTGTCAACAGGAGGAGGTGGATGTCTACCA 786  
DB 241 AlaCysHisProAlaGluAlaCysAlaGlyCysGlnGlnGluThrAspValValSer 260  
QY 787 GAGAGTCTTTGTGACATGTTGGTGGCGGAGCATTTGGACTGGACCAAAAATGACAC 846  
DB 261 GluSerProLeuSerAspThrGlySerGluAspValGlyThrGlyLeuLysAsnAlaAsn 280  
QY 847 AAATTCAGTGGCAAGAAAGCACCTAGGTGATTCGCTCCATTTGAGAAAGAAAGTGA 906  
DB 281 ArgLeuAsnArgGlnGluSerSerLeuLysSerLeuLysSerProPheGluLysGluSerGlu 300  
QY 907 CTGAGTCCCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 966  
DB 301 ProGluSerProMetAspValAspAsnSerLysAsnSerCysGlnAspSerGluAlaAsp 320  
QY 967 GAGAAACAGTCCAGTCTTTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1023  
DB 321 GluGluThrSerProGlyPheAspGluGlnGluAspSerSerSerAlaGlnThrAlaAsn 340  
QY 1024 AAATCTTCAAGTTGCCAAGCAAGAGAGCTGATGGGGGATCTTAGGAAACCGTATTGACT 1083

341	LysProSerArgPheGlnProArgGluAlaAspThrGluLeuArgLysArgSerSerAla	360
1084	AAGGGAAAGTGAAGTTAGATTGCATTTCCAAATTTCAA---GGAGAAAATAATGCTGGGACC	1140
361	LysGlyGlyGluIleArgLeuHisPheGlnPheGluGlyGlyGluSerArgAlaGlyMet	380
1141	AGTGACTTAATAGCCCAAGCCCATCTGGAAACTCTTCTAGCCTTAATAGTAGAGTGTAGAAGT	1200
381	AsnAspValAsnAlaLysArgProGlySerThrSerSerLeuAsnValGluCysArgAsn	400
1201	TCCAAGCAGCATCGAAAAAGGAGTCTTAAATATTCACAGTCAATTCATGAGATTTCCAG	1260
401	SerLysGlnHisGlyArgLysAspSerLysIleThrAspHisPheMetArgValProLys	420
1261	TCAGAGCAGACAGAAAAAGAACAAATGTGAAGTCAGACATCAAGAAACAGAAAGGAAGATT	1320
421	AlaGluAspLysArgLysGluGlnCysGluMetLysHisGlnArgThrGluArgLysIle	440
1321	CCAAAAATACATCCCACTTAACCTCCCTCCAGAGAAAGTGGCTGGAACTCTCTATTGAG	1380
441	ProLysTyrIleProProHisLeuSerProAspLysLysTrpLeuGlyThrProIleGlu	460
1381	GAATGAGAAAAATGCCCTCGGTGGGGATCCATTTGGCTTCCTTAAGACCATCTGCAAGT	1440
461	GluMetArgMetProArgCysGlyIleArgLeuProLeuArgProSerAlaAsn	480
1441	CACACAGTACTGTTCGGGTAGACCTTCTCAGACAGAGAGGTTCGGAACCTTTTCCA	1500
481	HisThrValThrIleArgValAspLeuLeuArgIleGlyGluValProLysProPhePro	500
1501	ACATTTACAAAGATTTGTGGGATAACAACATGTGAAATGCTTGTTCGGAAACAAAC	1560
501	ThrHisPheLysAspLeuTrpAspAsnLysHisValLysMetProCysSerGluGlnAsn	520
1561	TTGTACCTCTGGNAGATGAGATGGTGGCGAACTCGAGGAGTAGTGGGAGCTCAAT	1620
521	LeuTyrProValGluAspGluAsnGlyIleArgAlaAlaGlySerArgTrpGluLeuIle	540
1621	CAGACTGCATCTCTCAACAAATTCACACAGCCCCAGAACTTGAAGGATGGATTCGAAA	1680
541	GlnThrAlaLeuLeuAsnArgLeuThrArgProGlnAsnLeuLysAspAlaIleLeuLys	560
1681	TACAATGTGGCATATTCAAGAAATGGGACTTTACGTTTGGTTGATTTCTGGGATAAG	1740
561	TyrAsnValAlaTyrSerLysLysTrpAspPheThrAlaLeuIleAspPheTrpAspLys	580
1741	GTACTTGAAGNAGCAGAGGCCCAACATTTATATCAGTCCATTTTACCTGACATGGTGAAA	1800
581	ValLeuGluGluAlaGluAlaGlnHisLeuTyrGlnSerIleLeuProAspMetValLys	600
1801	ATTGCATCTGTCTGCCAAATATTTCACCCAGCCCAATACACATCCCTGGAACAGAAATG	1860
601	IleAlaLeuCysLeuProAsnIleCysThrGlnProIleProLeuLeuLysGlnLysMet	620
1861	AATCATTTCTGTACGATGTCCACAGAAACAGATCCCGAGTCTTTTAGCTAATGCTTCTTC	1920
621	AsnHisSerIleThrMetSerGlnGluGlnIleAlaSerLeuLeuAlaAsnAlaPhePhe	640
1921	TGCACATTTCCCGCAGGAATGCCAAGATGAATTCGGAGTATTCTAGTTACCCAGACATT	1980
641	CysThrPheProArgArgAsnAlaLysMetLysSerGluTyrSerSerTyrProAspIle	660
1981	AACTTCAATCGGTGTTTGAAGGACGTTTCATCAGGAAACACAGAAAAACTGAAACACTC	2040
661	AsnPheAsnArgLeuPheGluGlyArgSerSerArgLysProGluLysLeuLysThrLeu	680
2041	TTCTGTCTACTTTCGAAGAGTCACAGAAAAAAACCTACAGATTTGGTGCACATTTACAAGA	2100
681	PheCysTyrPheArgArgValThrGluLysLysProThrGlyLeuValThrPheThrArg	700
2101	CAGAGCTTGAAGATTTTCCAGAAATGGGAAGGTGTGAAAAGCTCTGACACGCTTACAC	2160
701	GlnSerLeuGluAspPheProGluTrpGluArgCysGluLysLeuLeuThrArgLeuHis	720

QY	2161	GTCACTTCAGGGGTACCA	PAGAAGGCACCGCGCGAGCATGCTACAGGTGGATTTTGCA	2222
Db	721	ValThrTyrGluGlyThrIleGluGlyAsnGlyGlnGlyMetLeuGlnValAspPheAla	740	
QY	2221	AATCGTTTTCTGGAGGTGTGTGACTGGTGGCGGACTTGTACAGAAGAAATCAGATTT	2280	
Db	741	AsnArgPheValGlyGlyGlyValThrSerAlaGlyLeuValGlnGluGluIleArgPhe	760	
QY	2281	TTAATCAATCCTCAATTTGATTTGTTTTCACGGCTGTTTCACTCAGGTGCTGGATCAATGAG	2340	
Db	761	LeuIleAsnProGluLeuIleValSerArgLeuPheThrGluValLeuAspHisAsnGlu	780	
QY	2341	TGCTTTATTATCACAGTACTGAACAGTACAGTGAATACACAGGCTATGCTGAAGACTTAT	2400	
Db	781	CysLeuIleIleThrGlyThrGluGlnTyrSerGluTyrThrGlyTyrAlaGluThrTyr	800	
QY	2401	CGTTGGCGCGAAGCCATGAAGTGGAGGAAAGAGCATTGGCAGCGCGCTGCACG	2460	
Db	801	ArgTrpAlaArgSerHisGluAspArgSerGluArgAspAspTrpGlnArgThrThr	820	
QY	2461	GAGATCGTTGCCATTGACGCACCTTCACTTCAGACGCTACCTCGCATCAGTTTGTGCTCGAG	2520	
Db	821	GluIleValAlaIleAspAlaLeuHisPheArgArgTyrLeuAspGlnPheValProGlu	840	
QY	2521	AAAGTGACGCTGAGCTTAACAGGCTTACTCGGATTCCTCCGCTCGTGGAGTTCCTTCT	2580	
Db	841	LysIleArgArgGluLeuAsnLysAlaTyrCysGlyPheLeuArgProGlyValSerSer	860	
QY	2581	GAATACTTTCTGCAGTGGCCACGGGAACTGGGGCTGTGGTGCCTTTGGGGGTGACGCT	2640	
Db	861	GluAsnLeuSerAlaValAlaThrGlyAsnTrpGlyCysGlyAlaPheGlyGlyAspAla	880	
QY	2641	AGATTAAGCCTTAATPACAGATCTCGACGCTGCTGGCGCTGAACGTGACGTGGTTTAT	2700	
Db	881	ArgLeuLysAlaLeuIleGlnIleLeuAlaAlaValAlaGluArgAspValValTyr	900	
QY	2701	TTCACTTTGGGACTCAGATTTGATGACAGACATTTACAGCATGCACCTTCCCTTACC	2760	
Db	901	PheThrPheGlyAspSerGluLeuMetArgAspIleTyrSerMetHisThrPheLeuThr	920	
QY	2761	GAGAGGAAGCTGATGTGTGAAAAGTGACAAAGTTATTGCTTAGATACTACAATGAAGA	2820	
Db	921	GluArgLysLeuThrValGlyGluValTyrLysLeuLeuLeuArgTyrTyrAsnGluGlu	940	
QY	2821	TGCAGAACTGTTCACCCCTGGACACAGACATCAAGCTTTATCATTATATACCATGCT	2880	
Db	941	CysArgAsnCysSerThrProGlyProAspIleLysLeuTyrProPheIleTyrHisAla	960	
QY	2881	GTTTCAGTCAGTGACAGACCACTGCATGCCAGGACAGAAAGCAGCG	2928	
Db	961	ValGluSerCysThrGlnThrThrAsnGlnProGlyGlnArgThrGly	976	
RESULT 12				
ABG72278				
ID	ABG72278	standard; protein; 977 AA.		
AC	ABG72278;			
XX				
XX				
DT	13-MAR-2003	(first entry)		
XX				
DE	Bovine poly (ADP-ribose) glycohydrolase (PARG) enzyme.			
XX				
KW	Bovine; poly (ADP-ribose) glycohydrolase; PARG; PARG expression;			
KW	cellular response; DNA damage; neoplastic disorder inducing agent;			
KW	oxidative stress; neoplastic disorder; myocardial infarction;			
KW	vascular stroke; neurodegenerative disorder; Alzheimer's disease;			
KW	Parkinson's disease; Huntington's disease; inborn genetic error;			
KW	reperfusion; ischaemia; aging; neurotoxicity; PARG activity; cytostatic;			
KW	neuroprotective; nootropic; antiparkinsonian; cardiant; vasotrophic;			
KW	anticonvulsant; cerebroprotective; enzyme.			
XX				
OS	Bos taurus.			

XX US2002132328-A1.  
 PN XX  
 PD 19-SEP-2002.  
 XX  
 XX 09-OCT-2001; 2001US-00973451.  
 PF XX  
 PR 01-MAY-1998; 98US-0083768P.  
 PR 30-APR-1999; 99US-00302812.  
 XX  
 XX (JACO/) JACOBSON M K.  
 PA (JACO/) JACOBSON E L.  
 PA (AMEJ/) AME J.  
 PA (LINW/) LIN W.  
 XX  
 PI Jacobson MK, Jacobson EL, Ame J, Lin W;  
 XX WPI; 2003-155895/15.  
 DR N-PSDB; ABX14477.  
 XX  
 XX New nucleic acid molecule encoding a polypeptide with poly(ADP-ribose)  
 PT glycohydrolase activity, for preventing, treating, or ameliorating a  
 PT disease condition, e.g. neoplastic disorder, myocardial infarction or  
 PT vascular stroke.  
 XX  
 XX Claim 28; Fig 16; 86pp; English.  
 PS  
 XX The present invention relates to the isolation of poly(ADP-ribose)  
 CC glycohydrolase (PARG) from several species, and the polynucleotide  
 CC sequences encoding them. Methods for inhibiting PARG expression or  
 CC overexpressing PARG are also disclosed. PARG is involved in the cellular  
 CC response to DNA damage, and is associated with the body's response to  
 CC neoplastic disorder inducing agents and oxidative stress. The  
 CC polynucleotide sequences encoding PARG and PARG modulators are useful for  
 CC preventing, treating, or ameliorating diseases such as neoplastic  
 CC disorders, myocardial infarction, vascular stroke, neurodegenerative  
 CC disorders (e.g. Alzheimer's disease, Parkinson's disease or Huntington's  
 CC disease), inborn genetic errors, reperfusion following ischaemia, aging,  
 CC and neurotoxicity. The polynucleotide sequences are also useful in gene  
 CC therapy. The methods are useful for identifying an agent that modulates  
 CC PARG activity, identifying a mutant PARG allele in an individual, or  
 CC screening candidate molecules for PARG modulating activity. The present  
 CC sequence represents bovine PARG enzyme  
 XX  
 SQ Sequence 977 AA;

Alignment Scores:  
 Pred. No.: 0 Length: 977  
 Score: 4332.50 Matches: 820  
 Percent Similarity: 89.86% Conservative: 57  
 Best Local Similarity: 84.02% Mismatches: 90  
 Query Match: 63.97% Indels: 9  
 DB: 6 Gaps: 5

US-09-302-812-5 (1-3814) x ABG72278 (1-977)

QY 28 ATGAGTGGGCGCCGGCTGGAGCCCTGCACGAAA---GGCGCTGGGGCGCGCTGGA 84  
 Db 1 MetSerAlaGlyProGlyCysGluProCysThrIysArgProArgTTPAspAlaAla 20  
 QY 85 ACTTCTGGCGCATGCTCTCGGATCCCGGAGCTTCCTCGGACGCGAGCGGTTC 144  
 Db 21 ThrSerProProAlaAlaSerAspAlaArgSerPheProGlyArgGlnArgValLeu 40  
 QY 145 GACCCAGGAGCTCCGCTCCAGTTCAGGCTCCCTCGCTCCGACGCTGCTGCTCG 204  
 Db 41 AspSerIysAspAlaProValGlnPheArgValProProSerSerSerGlyCysAlaLeu 60  
 QY 205 GGGCGGCGGACCGCAGAGGCAACGCCACCTCGTTGTTTCAACAAAGAACTATT 264  
 Db 61 GlyArgAlaGlyGlnHisArgGlySerAlaThrSerLeuValPheLysGlnLysThrIle 80  
 QY 265 ACTACTGGATGGATATAAAGGACCAAGACAGCTGAATCAGAA-----AGTAA 315

Db 81 ThrSerTrpMetAspThrIysGlyIleIysThrValGluSerGluSerLeuHisSerIys 100  
 QY 316 GAAACAAACATACAGAAATTGACTCCATGATGAGTCTGTGCAGAAAAGATACTTTTAC 375  
 Db 101 GluAsnAsnThrArgGluGluSerMetSerValGlnLysAspAsnPheTyr 120  
 QY 376 CCACATAAGGTGGAAAATTCGAAAATCTTCTCAGCTAAATCTTGATAAATCACCACA 435  
 Db 121 GlnHisAsnMetGluLysLeuGluAsnValSerGlnLeuGlyPheAspLysSerProVal 140  
 QY 436 GAAAAGACTTCACAGATATTTGAACCAACAGACAGACTGCAGTGTCTGCAAGTGGCAGAT 495  
 Db 141 GluLysGlyThrGlnTyrLeuLysGlnHisGlnThrAlaIaMetCysLysTrpGlnAsn 160  
 QY 496 GAAGGAAAGCATGCAGAACAGCTTTTGGCAAGTGAAGCTCCCGGGGAGCTCCGCTACCA 555  
 Db 161 GluGlyProHisSerGluArgLeuGluSerGluProProAlaValThrLeuValPro 180  
 QY 556 AAGCAGCTTACTAATGCTTAACATTTGGTCAGTCACCCACACTGATGACACAGTGCACA 615  
 Db 181 GluGlnPheSerAsnAlaAsnValAspGlnSerSerProLysAspAspHisSerAspThr 200  
 QY 616 GATCATGAAGAGACAGACACAATCAGCAGTTTCTTACACCTATATAAACTTGCATAATACA 675  
 Db 201 AsnSerGluGluSerArgAspAsnGlnGlnPheLeuThrHisValLysLeuAlaAsnAla 220  
 QY 676 AAGCCAACTAGTAGAGATGGGCAG-----GCCAGAACAACTGTAAGTGCAGTGA 726  
 Db 221 LysGlnThrMetGluAspGluGlnGlyArgGluAlaArgSerHisGlnLysCysGlyLys 240  
 QY 727 TCTCGCCAGCTGTGAAAGACTGTACAGGCTGTCAACAGGAGGAGGTGGATGCTACCA 786  
 Db 241 AlaCysHisProAlaGluAlaCysAlaGlyCysGlnGlnGluThrAspValValSer 260  
 QY 787 GAGAGTCTTTTGTGATGTTGGTGGCCGAGACACTGGAACTGCACCAAAAAATACACAAC 846  
 Db 261 GluSerProLeuSerAspThrGlySerGluAspValGlyThrGlyLeuLysAsnAlaAsn 280  
 QY 847 AAATTGCTGACAAAGAAAGACAGCTAGTGTATTCGCTCCATTTGAGAAAGAAAGTGA 906  
 Db 281 ArgLeuAsnArgGlnGluSerSerLeuGlyAsnSerProProPheGluLysGluSerGlu 300  
 QY 907 CCTGAGTCACCAATGGATGTAGCAACTCGAGAACACAGTGTCAAGATTTCAGAGCAGAT 966  
 Db 301 ProGluSerProMetAspValAspAsnSerLysAsnSerCysGlnAspSerGluAlaAsp 320  
 QY 967 GAAGAAACAAAGTCCAGTCTTTTGATGACAAAGATGATCGTTCTCTCC---CAACAGCAAA 1023  
 Db 321 GluGluThrSerProGlyPheAspGluGlnGluAspSerSerSerAlaGlnThrAlaAsn 340  
 QY 1024 AACTTTCAAGTTGCCAAGCAAGAGCTGATGGCGATCTTAGGAACGGTATTTCACT 1083  
 Db 341 LysProSerArgPheGlnProArgGluAlaAspThrGluLeuArgLysArgSerSerAla 360  
 QY 1084 AAGGAAAGTCAAGTAGATTGCAATTTCCAATTTGAA---GGAGAAAATAATCTGGCGACC 1140  
 Db 361 LysGlyGlyGluIleArgLeuHisPheGlnPheGluGlyGlyGluSerArgAlaGlyMet 380  
 QY 1141 AGTGACTTAATATGCAAGCCATCTGGAAACTCTTCTAGCCCTTAATGTAGAGTGTAGAAGT 1200  
 Db 381 AsnAspValAsnAlaLysArgProGlySerThrSerSerLeuAsnValGluCysArgAsn 400  
 QY 1201 TCCACGACATCGAAAAGGGATCTTAATATTACAGATCATTTTCATGAGAAATTCACAG 1260  
 Db 401 SerLysGlnHisGlyArgLysAspSerLysIleThrAspHisPheMetArgValProLys 420  
 QY 1261 TCAGAGACACAGAAAAGAAACAATGTGAAGTCAGACATCAAGAACAGAAAAGAAAGATT 1320  
 Db 421 AlaGluAspLysArgLysGluGlnCysGluMetLysHisGlnArgThrGluArgLysIle 440  
 QY 1321 CCAAAATACATCCCACTAACCTCCCTCCAGAGAGAGAGTGGCTGGGAATCTCTATTGAG 1380

Db 441 ProLysTyrIleProHisLeuSerProAspLysLysTrpLeuGlyThrProIleGlu 460  
QY 1381 GAAATGAGAAAATGCTCGGTGGGATCCATTGGCTTCCTTAAGACCATCTGCAAGT 1440  
Db 461 GluMetArgArgMetProArgCysGlyIleArgLeuProProLeuArgProSerAlaAsn 480  
QY 1441 CACACAGTACTGTTCGGGTAGAGCTTCTCAGAGCAGAGAGGTTCGGAACCTTTTCCA 1500  
Db 481 HisThrValThrIleArgValAspLeuLeuLeuArgIleGlyGluValProLysProPhePro 500  
QY 1501 ACAATTACAAAGATTGTGGGATAACAAACATGTGAAATGCTTGTTCGGACAAAC 1560  
Db 501 ThrHisPheLysAspLeuTrpAspAsnLysHisValLysMetProCysSerGluGlnAsn 520  
QY 1561 TTCTACCTCTGGAAGTGAAGTGTGAGCGACCTGAGCGAGTGTGAGCGAGTGTGAGCTATT 1620  
Db 521 LeuTyrProValGluAspGluAsnGlyGluArgAlaAlaGlySerArgTrpGluLeuIle 540  
QY 1621 CAGACTGCCTTCTCAACAAATTCACAGACCCAGAACCTTGAAGATGCGATTCTGAAA 1680  
Db 541 GlnThrAlaLeuLeuAsnArgLeuThrArgProGluAsnLeuLysAspAlaIleLeuLys 560  
QY 1681 TACAATGTGCATATTCTTAAGAAATCGGACTTTACAGCTTTGTTGATTCTCGGATAAG 1740  
Db 561 TyrAsnValAlaTyrSerLysLysTrpAspPheThrAlaLeuIleAspPheTrpAspLys 580  
QY 1741 GTACTTGAAGACGACAGCCCAACATTTATATCATCTCCATTTTACCTGACATGGTGAAA 1800  
Db 581 ValLeuGluGluAlaGluAlaGlnHisLeuTyrGlnSerIleLeuProAspMetValLys 600  
QY 1801 ATTGCACTCTGTCTGCCAAATATTTCACCCAGCCCAATACCACTCTCTGAAACAGAGATG 1860  
Db 601 IleAlaLeuCysLeuProAsnIleCysThrGlnProIleProLeuLeuLysGlnLysMet 620  
QY 1861 AATCATCTGTGCAGATGTCACAGAACAGATCGCCAGTCTTTTACGTAATGCTTTCTTC 1920  
Db 621 AsnHisSerIleThrMetSerGlnGluGlnIleAlaSerLeuLeuAlaAsnAlaPhePhe 640  
QY 1921 TGCACATTTCCCGACGGAATGCCAGATGCAATCGGAGTATTCAGTACCTACCCAGACATT 1980  
Db 641 CysThrPheProArgArgAsnAlaLysMetLysSerGluTyrSerSerTyrProAspIle 660  
QY 1981 AACTTCAATCGGTGTTCGAAGACCTTCATCAAGGAAACAGAAAACCTGAAACACATC 2040  
Db 661 AsnPheAsnArgLeuPheGluGlyArgSerSerArgLysProGluLysLeuLysThrLeu 580  
QY 2041 TTCTGTACTTTGGAAGATGTCAGAGAAAACCTTACAGGATGTGTGACATTTACAGA 2100  
Db 681 PheCysTyrPheArgArgValThrGluLysLysProThrGlyLeuValThrPheThrArg 700  
QY 2101 CAGAGTCTTGAAGATTTTCCAGAATGGGAAGGTGTGAAAGCCTCTGACAGCTTACAC 2160  
Db 701 GlnSerLeuGluAspPheProGluTrpGluArgCysGluLysLeuLeuThrArgLeuHis 720  
QY 2161 GTCACTACAGGGTACCATAGAAGCAACGCGCCAGGATGCTACAGGTGATTTTGA 2220  
Db 721 ValThrTyrGluGlyThrIleGluGlyAsnGlyGlnGlyMetLeuGlnValAspPheAla 740  
QY 2221 AATCGTTTGTGGAGTGTGTGATGCTGTGGGACTTGTACAGAAGAAATCAGATT 2280  
Db 741 AsnArgPheValGlyGlyGlyValThrSerAlaGlyLeuValGlnGluIleArgPhe 760  
QY 2281 TTAATCAATCCTGAATTGATTGTTTCACGCTGTTCACCTGAGTGTGATCAATGAG 2340  
Db 761 LeuIleAsnProGluLeuIleValSerArgLeuPheThrGluValLeuAspHisAsnGlu 780  
QY 2341 TGTCTATTATCACAGGTACTCAACAGTACAGTGAATACACAGGCTATGCTCAAACTTAT 2400  
Db 781 CysLeuIleIleThrGlyThrGluGlnTyrSerGluTyrThrGlyThrAlaGluThrTyr 800  
QY 2401 CGTTGGGCCGACGATGAGTGGAGTGAAGACCATTTGGCAGCGCGCTGCACG 2460  
Db 801 ArgTrpAlaArgSerHisGluAspArgSerGluArgAspTrpGlnArgArgThrThr 820

QY 2461 GAGATCGTTGCCATTCAGCGCACTTCACTTCAGAGCTACCTCGATCAGTTTGCCTGAG 2520  
Db 821 GluIleValAlaIleAspAlaLeuHisPheArgTyrIleuAspGlnPheValProGlu 840  
QY 2521 AAAGTCAGACGCTGAGCTTAACAGGCTTACTCGGGATTCTCGTCTCGTGGAGTTCCTTCT 2580  
Db 841 LysIleArgArgGluLeuAsnLysAlaTyrCysGlyPheLeuArgProGlyValSerSer 860  
QY 2581 GAAATCTTCTGCACTGCGCCAGCGGAAACTGGGGCTGTGGTCCCTTTGGGGGTGAGCT 2640  
Db 861 GluAsnLeuSerAlaValAlaThrGlyAsnTrpGlyCysGlyAlaPheGlyGlyAspAla 880  
QY 2641 AGATTAAAGCCCTTAATACAGATCTCGGAGCTGCTGCGGCTGAACTGACGTGCTGTTAT 2700  
Db 881 ArgLeuLysAlaLeuIleGlnIleLeuAlaAlaValAlaGluArgAspValValTyr 900  
QY 2701 TTCACCTTTGGGACTCAGAGTTGATGAGAGCATTTACAGCATGACACACTTTCCCTTACC 2760  
Db 901 PheThrPheGlyAspSerGluLeuMetArgAspIleTyrSerMetHisThrPheLeuThr 920  
QY 2761 GAGAGAACTGGATGTTGGAAAAGTGTACAAGTTATTGCTTAGATACATACATGAGAA 2820  
Db 921 GluArgLysLeuThrValGlyGluValTyrLysLeuLeuLeuArgTyrTyrAsnGluGlu 940  
QY 2821 TGCAGAACTGTTCACCCCTGGACCCAGACATCAAGCTTTATCCATTATATACCATGCT 2880  
Db 941 CysArgAsnCysSerThrProGlyProAspIleLysLeuTyrProPheIleTyrHisAla 960  
QY 2881 GTTGAAGTCAAGTGCAGACCACTGACATGCCAGGACAGACAGAGCAGGC 2928  
Db 961 ValGluSerCysThrGlnThrThrAsnGlnProGlyGlnArgThrGly 976  
RESULT 13  
ABB59491  
ID ABB59491 standard; protein; 768 AA.  
AC ABB59491;  
XX 26-MAR-2002 (first entry)  
XX Drosophila melanogaster polypeptide SEQ ID NO 5265.  
XX Drosophila; developmental biology; cell signalling; insecticide;  
XX pharmacological.  
XX Drosophila melanogaster.  
XX WO200171042-A2.  
XX 27-SEP-2001.  
XX 23-MAR-2001; 2001WO-US009231.  
XX 23-MAR-2000; 2000US-0191637P.  
XX 11-JUL-2000; 2000US-00614150.  
XX (PEKE ) PE CORP NY.  
XX Venter JC, Adams M, Li PWD, Myers EW;  
PI WPI: 2001-656860/75.  
XX N-PSDB; ABL03594.  
XX New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signaling and cell-cell  
PT interactions.  
XX Disclosure; SEQ ID NO 5265; 21pp + Sequence Listing; English.  
XX The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and

cell-cell interactions in higher eukaryotes for the development of insecticides, therapeutics and pharmaceutical drugs. The invention discloses genomic DNA sequences (ABU16176-ABU30511), expressed DNA sequences (ABU01840-ABU16175) and the encoded proteins (ABB57737-ABB72072). The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences

XX SQ Sequence 768 AA;

Alignment Scores:

Pred. No.:	1-72e-93	Length:	768
Score:	1074.50	Matches:	244
Percent Similarity:	58.12%	Conservative:	96
Best Local Similarity:	41.71%	Mismatches:	194
Query Match:	15.86%	Indels:	51
DB:	4	Gaps:	15

US-09-302-812-5 (1-3814) x ABB59491 (1-768)

QY 1249 AGAATTCACAGTACAGAGACAGAGAAAGAACATGTGAAGTCAGATCAAGAACA 1308  
 DB 45 ArgMetSerLysSerProAspGlyGlyIleSerGluIleGluThrGluGlu 62  
 QY 1309 GAAAGAGATTCACAAATACATCCACCTAACCTCCCTCCA-----GAGAAGAAGTGG 1362  
 DB 63 -----ProGluAsnLeuAlaAsnSerLeuAspSerTyr 74  
 QY 1363 CTGGGAATCTTATTCAGGAAATGAGAAATGCTCGGTGGATTCATTTGCTTCC 1422  
 DB 75 ArgGlyValSerMetGluAlaIleHisArg---AsnArgGlnProPheGluLeuGluAsn 93  
 QY 1423 TTAAGACCA---TCTGCAAGTCACAGTACAGTGTTCGGGTAGACCTTCAGAGCAGGA 1479  
 DB 94 LeuProProValThraAlaGlnLeuHisArgValMetTyrGlnLeuProIleArgGlu 113  
 QY 1480 GAGGTTCCGAACCTTTTCCACACATTACAAAGATTTCTGGGATAACAAACATGTGAA 1539  
 DB 114 ThrProProArgProTyrLysSerProGlyLys-----TyrAspSerGluHisValArg 131  
 QY 1540 ATGCTTTGTCGGAACAAACTGTACCTGTGGAGATGAGAATGGTGAGCGAACTGCA 1599  
 DB 132 LeuProCysAlaProGluSerLysTyrProArgGluAsnProAspGlySerThrThrIle 151  
 QY 1600 GGGAGTAGTGAGCTCATTCAGACTGCATCTCTCAACAAATTCACAGCCCCAGAAC 1659  
 DB 152 AspPheArgTyrGluMetIleGluArgAlaLeuLeuGlnProIleLysThrCysGluGlu 171  
 QY 1660 TTGAAGTAGCGATTTCTGAATACATGTGGCATATTTCTAAGAAATGGGACTTTACAGCT 1719  
 DB 172 LeuGlnAlaAlaIleSerTyrAsnThrThrTyrArgAspGlnTyrHisPheArgAla 191  
 QY 1720 TTGGTTGATTTCTGGATAAGTACTTCAGAGCAGAGGCCCAACATTTATACAGTCC 1779  
 DB 192 LeuHisGlnLeuLeuAspGluGluLeuAspGluSerGluThrArgValPhePheGluAsp 211  
 QY 1780 ATTTTACCTGCATGCTGAATATTCATCTCTGTCGCAAAATATTTCACCCAGCAATA 1839  
 DB 212 LeuLeuProArgIleIleArgLeuAlaLeuArgLeuProAspGluIleGlnSerProVal 231  
 QY 1840 CCATCTCTGAAACAGAGAATGAATCATTTCTGTCAGCATGTCTACAGAAACAGATCCCACT 1899  
 DB 232 ProLeuLeuLysHisHisLysAsnAlaSerLeuSerLeuSerGlnGlnGlnIleSerCys 251  
 QY 1900 CTTTGTAGTAAATGCTTTCTCTGCAATTTCCCGACCGAATGCC---AAGATGAATTCG 1956  
 DB 252 LeuLeuAlaAsnAlaPheLeuCysThrPheProArgArgAsnThrLeuLysArgLysSer 271  
 QY 1957 GAGTATTCAGTACCCAGACATTAATTCATTCATTCGTTTGAAGGAGCTTCATCAAGG 2016  
 DB 272 GlufySerThrProProAspIleAsnPheAsnArgLeuTyrGlnSerThrGlyProAla 291  
 QY 2017 AAACACAGAAAACTGAAACACACTCTCTGCTACTTCTCGAAGAGTC-----ACAGAGAA 2070

Db 292 ValLeuGluLysLeuLysCysIleMethIstYrPheArgArgValCysProThrGluArg 311  
 QY 2071 AAA-----CCTACAGATTGGTGACATTTACAGACAGAGT-----CTTGAA 2112  
 Db 312 AspAlaSerAsnValProThrGlyValThrPheValArgSerGlyLeuProGlu 331  
 QY 2113 GATTTTCCAGATGGAAAGGTGTGAAAAGCCTCTG-----ACAGCTTACAGCTCACT 2166  
 Db 332 HisLeuIleAspTyrSerGlnSerAlaAlaProLeuGlyAspValProLeuHisValAsp 351  
 QY 2167 TACGAGGTACCATAGAAAGCAAGCCGAGCATCTACAGTGGATTTTCCAAATCCT 2226  
 Db 352 AlaGluGlyThrIleGluAspGluGlyIleGlyLeuLeuGlnValAspPheAlaAsnLys 371  
 QY 2227 TTTGTTGAGGTGTGTGACTGCTGGGACCTGTGTACAGAAAGAAATCAGATTTTAAATC 2286  
 Db 372 TyrLeuGlyGlyValLeuGlyHisGlyCysValGlnGluIleArgPheValIle 391  
 QY 2287 AATCTGATTCATTTTACGGCTGTTCATCTGAGGTGCTGGATCACAATGAGTGTCTT 2346  
 Db 392 CysProGluLeuLeuValGlyLysLeuPheThrGluCysLeuArgProPheGluAlaLeu 411  
 QY 2347 ATTATCACAGGTACTGAAGTGGAGTGAAGAGCAAGTATGGCAGCGGCTGACGAGATC 2406  
 Db 412 ValMetLeuGlyAlaGluArgTyrSerAsnTyrThrGlyTyrAlaGlySerPheGluTyr 431  
 QY 2407 GCGCAAGCCATGAAGTGGAGTGAAGAGCAAGTATGGCAGCGGCTGACGAGATC 2466  
 Db 432 SerGlyAsnPheGluAspSerThrProArgAspSerSerGlyArgGlnThrAlaIle 451  
 QY 2467 GTTGCCATTCAGCTTCTACCTGACGCTACCTGATCAGTACCTGATGCTGCTGCTGAGAAAGT 2526  
 Db 452 ValAlaIleAspAlaLeuHisPheAlaGlnSerHisGlnTyrArgGluAspLeuMet 471  
 QY 2527 AGACTGTAGCTTAAACAGCTTACTGCGGATTCCTCCGT-----CCT 2568  
 Db 472 GluArgGluLeuAsnLysAlaIleTyrIleGlyPheValHisTyrMetValThrProPro 491  
 QY 2569 GGAGTTCCTCTCGAAAAATCTTTCTGAGTGGCCAGCGGAAACTGGGGCTGTGGTCCCTTT 2628  
 Db 492 Gly-----ValAlaThrGlyAsnTyrGlyCysGlyAlaPhe 503  
 QY 2629 GGGGTGACGTAGATTAAGACCTTAATACAGATCCTCGGACGCTGCTGCGCTGAAAGT 2688  
 Db 504 GlyGlyAspSerTyrLeuLysAlaLeuLeuGlnMetValCysAlaGlnLeuGlyArg 523  
 QY 2689 GACGTGGTTTATTTACCTTTGGGACATCAGAGTTGATGACAGACATTTACAGCATGCAC 2748  
 Db 524 ProLeuAlaTyrTyrThrPheGlyAsnValGluPheArgAspPheHisGluMetTyr 543  
 QY 2749 ACTTCTCTTACCGAGAGAGTGTGGAAAGTGTACAGTATTCCTGATAGTAC 2808  
 Db 544 LeuLeuPheArgAsnAspGlyThrThrValGlnGlnLeuTyrSer---IleLeuArgSer 562  
 QY 2809 TACATGAAGATGACAGAACTGTTCCACCCCTCGACCA-----GAC 2850  
 Db 563 TyrSerArgLeuIleLysGluLysSerSerLysGluProArgGluAsnLysAlaSerLys 582  
 QY 2851 ATCAAGCTTTATCCATTCATATACCTGCTGTCAGTCAAGTCAGTCAGACACCTGACATG 2910  
 Db 583 LysLysLeuTyrAspPheIle-----LysGluGluLeuLysLysValArgAspVal 599  
 QY 2911 CCAGCAGAGAGGCA 2925  
 Db 600 ProGlyGluGlyAla 604

RESULT 14  
 AAE25632  
 ID AAE25632 standard; protein; 768 AA.  
 XX  
 AC AAE25632;  
 XX





```
Db 232 ProLeuLeuLysHisHisLysAsnAlaSerLeuSerLeuSerGlnGlnGlnIleSerCys 251
QY 1900 CTTTGTAGCTAATGCTTCTTCTGCACATTCCCGGACGGAATGCC---AAGATGAATCG 1956
Db 252 LeuLeuAlaAsnAlaPheLeuCysThrPheProArgAsnThrLeuLysArgLysSer 271
QY 1957 GAGTATTCTAGTTACCCAGACATTAATCTCAATCGGTTGTTGAAGGAGCTTCATCAAGG 2016
Db 272 GluTyrSerThrPheProAspIleAsnPheAsnArgLeuTyrGlnSerThrGlyProAla 291
QY 2017 AACCCAGAAAACCTGAACACACTCTCTGCTACTCTTCGAGAGTC-----ACAGAGAAA 2070
Db 292 ValLeuGluLysLeuLysCysIleMetHisTyrPheArgArgValCysProThrGluArg 311
QY 2071 AAA-----CCTACAGGATTGGTGACATTTTACAGACAGAGT-----CTTGAA 2112
Db 312 AspAlaSerAsnValProThrGlyValValThrPheValArgSerGlyLeuProGlu 331
QY 2113 GATTTTCCAGATGGGAAGGTGTGAAGCCCTCTG-----ACAGCTTACACGTCAC 2166
Db 332 HisLeuIleAspTrpSerGlnSerAlaAlaProLeuGlyAspValProLeuHisValAsp 351
QY 2167 TAGCGGGTACCATGAAGCAAGCGCGAGGCGATGCTACAGGTGGATTTTGCAAAATCGT 2226
Db 352 AlaGluGlyThrIleGluAspGluGlyIleGlyLeuLeuGlnValAspPheAlaAsnLys 371
QY 2227 TTTGTTGGAGGTGCTGTGATGCTGGCGGACTTGTACAGAAGAAATCAGATTTTAAATC 2286
Db 372 TyrLeuGlyGlyGlyValLeuGlyHisGlyCysValGlnGluIleArgPheValIle 391
QY 2287 AATCCTGAATTGTTGTTTACCGGCTGTGTTACAGTGTGCTGATCACAATGAGTGTCTT 2346
Db 392 CysProGluLeuLeuValGlyLysLeuPheThrGluCysLeuArgProPheGluAlaLeu 411
QY 2347 ATTATCACAGGTACTGAACAGTACAGTGAATACACAGGCTATGCTGAAACTTATCGTTGG 2406
Db 412 ValMetLeuGlyAlaGluArgTyrSerAsnTyrThrGlyTyrZalaglySerPheGluTrp 431
QY 2407 GCCCGAAGCCATGAAGATGGAGTGAAGAGCGATTGGCAGCGGCTGCACGAGATC 2466
Db 432 SerGlyAsnPheGluAspSerThrProArgAspSerSerGlyArgArgGlnThrAlaIle 451
QY 2467 GTTGCATTCACGCACCTTCACTTCAGACGCTCCTCGATCAGTTCGTGCTGAGAAAGTG 2526
Db 452 ValAlaIleAspAlaLeuHisPheAlaGlnSerHisHisGlnTyrArgGluAspLeuMet 471
QY 2527 AGACGTGAGCTTAACAGGCTTACTCGGATTCCTCGT-----CCT 2568
Db 472 GluArgGluLeuAsnLysAlaTyrIleGlyPheValHisTrpMetValThrProProPio 491
QY 2569 GGAGTTCTCTTGAAATCTTCTGACAGTGGCCACGGGAAACTGGGCTGTGTGCTTT 2628
Db 492 Gly-----ValAlaThrGlyAsnTrpGlyCysGlyAlaPhe 503
QY 2629 GGGGGTGACCTAGATTAAAGCCTTAATACAGATCCTGCAGCTGCTGCGGCTGACGT 2688
Db 504 GlyGlyAspSerTyrLeuLysAlaLeuLeuGlnLeuMetValCysAlaGlnLeuGlyArg 523
QY 2689 GACGTGGTTTATTTACCTTTTGGGACTCAGAGTTGATGAGAGACATTTACAGCATGCAC 2748
Db 524 ProLeuAlaTyrTyrThrPheGlyAsnValGluPheArgAspAspPheHisGluMetTrp 543
QY 2749 ACTTTCCTTACCCAGAGGAAGCTGGATGTTGGAAAGTGTACAAAGTTATTGCTTAGATAC 2808
Db 544 LeuLeuPheArgAsnAspGlyThrThrValGlnGlnLeuTrpSer---IleLeuArgSer 562
QY 2809 TACATGAAGATGCAGAAACTGTTCCACCCCTGGACCA-----GAC 2850
Db 563 TyrSerArgLeuIleLysGluLysSerSerLysGluProArgGluAsnLysAlaSerLys 582
QY 2851 ATCAAGCTTTTATCCATTATACATACCATGCTGTTGAGTCAAGTCAAGTGCAGACCATG 2910
```

```
Db 583 LysLysLeuTyrAspPheIle-----LysGluGluLeuLysLysValArgAspVal 599
QY 2911 CCAGGACAGAGGCA 2925
Db 600 ProGlyGluGlyAla 604
```

Search completed: May 26, 2004, 15:49:22  
Job time : 274.747 secs









Db 301 AspAenSerArgAenSerCysGlnAspSerGluAlaAspGluThrSerProValPhe 320  
QY 988 GATGAGCAAGATGATCGTTCTCCCAACAGCAATAAATCTTCAAGTTGCGCAAGCAAGA 1047  
Db 321 AspGluGlnAspAspArgSerSerGlnThrAlaAenLysLeuSerSerCysGlnAlaArg 340  
QY 1048 GAAGCTGATGGCGATCTTCTAGGAAACGGTATTGACTAAGGGAAGTGAAGTTAGATTGCAT 1107  
Db 341 GluAlaAspGlyAspLeuArgLysArgTyrLeuThrLysGlySerGluValArgLeuHis 360  
QY 1108 TTCCAAATTTGAGGAGAAATAATCTCGGACCGAGTGAATTAATGCCAAGCCATCTCGA 1167  
Db 361 PheGlnPheGluGlyGluAsnAlaGlyThrSerAspLeuAsnAlaLysProSerGly 380  
QY 1168 AACTCTTCTAGCCTTAATGTAGAGTGTAGAGTTCCAAAGCAGCATGGAAAGAGGATTCT 1227  
Db 381 AenSerSerSerLeuAsnValGluCysArgSerSerLysGlnHisGlyLysArgAspSer 400  
QY 1228 AAAATTACAGATCATTTTCATGAGAAATTTCCAAGTCAGAGGACAGAGAAGAAACAATGT 1287  
Db 401 LysIleThrAspHisPheMetArgIleSerLysSerGluAspArgArgLysGluGlnCys 420  
QY 1288 GAAGTCAGACATCAAGACAGAAAGAGAGATTCCAAATATACATCCCACTTAACCTCCCT 1347  
Db 421 GluValArgHisGlnArgThrGluArgLysIleProLysTyrIleProProAsnLeuPro 440  
QY 1348 CCAGAGAAGAGTGGCTGGGAATCCTATTGAGGAAATGAGAAATGCCCTGGTGGG 1407  
Db 441 ProGluLysLysTyrLeuGlyThrProIleGluMetArgLysMetProArgCysGly 460  
QY 1408 ATCCATTTGCCCTTCTTAAGACCATCTGCAAGTCACACAGTGCATGTTGGGTAGACCTT 1467  
Db 461 IleHisLeuProSerLeuArgProSerAlaSerHisThrValThrValArgValAspLeu 480  
QY 1468 CTGAGCAGCAGAGGTTCCGAACTTTCCACACATTAACAAGATTGTTGGGATAAC 1527  
Db 481 LeuArgAlaGlyGluValProLysProPheProThrHisTyrLysAspLeuTyrAspAsn 500  
QY 1528 AAAATGTGAAATTCCTTTGTCGAAACAAACTTGTACCTCTGGAAGATGAGAATCGT 1587  
Db 501 LysHisValLysMetProCysSerGluGlnAsnLeuTyrProValGluAspGluAsnGly 520  
QY 1588 GAGCGAATCGCAGGAGTAGTGGAGCTCATTCAGACTGCACCTTCTCAACAATTCACA 1647  
Db 521 GluArgThrAlaGlySerArgTyrGluLeuIleGlnThrAlaLeuLeuAsnLysPheThr 540  
QY 1648 CGACCCAGACTTCAAGATGCGATTTCTGGAATACATGTCGCAATATTCTAAGAAATGG 1707  
Db 541 ArgProGlnAsnLeuLysAspAlaIleLeuLysTyrAsnValAlaTyrSerLysLysTyr 560  
QY 1708 GACTTTACAGCTTTGTTGATTTCTGGGATAGGTACTTGAAGAAGCAGAGCGCCCAACAT 1767  
Db 561 AspPheThrAlaLeuValAspPheTyrAspLysValLeuGluGluAlaGlnHis 580  
QY 1768 TTATATAGTCCATTTTACCTGACATGTGTGAATGCACTCTGTCTGCCAAATATTTCG 1827  
Db 581 LeuTyrGlnSerIleLeuProAspMetValLysIleAlaLeuCysLeuProAsnIleCys 600  
QY 1828 ACCGAGCAATACCACTCTGAAACAGAGATGATCATTTCTGTCACATTTCCCGCAGCAAG 1887  
Db 601 ThrGlnProIleProLeuLeuLysGlnLysMetAsnHisSerValThrMetSerGlnGlu 620  
QY 1888 CAGATCGCAGTCTTTTGTAGTAAATGCTTTCTGTCACATTTCCCGCAGCAAG 1947  
Db 621 GlnIleAlaSerLeuLeuAlaAsnAlaPhePheCysThrPheProArgArgAsnAlaLys 640  
QY 1948 ATGAATCGAGTATTTCTAGTTACCCAGACATTAATCAATCGTGTGTTTGAAGAGCGT 2007  
Db 641 MetLysSerGluTyrSerSerTyrProAspIleAsnPheAsnArgLeuPheGluGlyArg 660  
QY 2008 TCATCAAGGAACACAGAAACTGAAACACCTCTTCTGCTACTTTCTGGAAGATCACAG 2067  
Db 661 SerSerArgLysProGluLysLeuLysThrLeuPheCysTyrPheArgArgValThrGlu 680

QY 2068 AAAAAACCTACAGGATTTGTCACATTTACAGACAGAGCTTTGAAGATTTTCCAGATGG 2127  
Db 681 LysLysProThrGlyLeuValThrPheThrArgGlnSerLeuGluAspPheProGluTyr 700  
QY 2128 GAAAGGTGTGAAAGCCCTCTGACACGCTTACACGCTTACACGCTTACAGAGGTACCATAGAAGC 2187  
Db 701 GluArgCysGluLysProLeuThrArgLeuHisValThrTyrGluGlyThrIleGluGly 720  
QY 2188 AACGCGCAGGACGCTACAGCTGATTTTCGCAATCGTTTGTGGAGGTGGTGTGACT 2247  
Db 721 AsnGlyArgGlyMetLeuGlnValAspPheAlaAsnArgPheValGlyGlyValThr 740  
QY 2248 GGTGCGGACTTGTACAGAAGAAATCAGATTTTAAATCAATCTCTGCAATGATTTTCA 2307  
Db 741 GlyAlaGlyLeuValGlnGluLeuIleArgPheLeuIleAsnProGluLeuIleValSer 760  
QY 2308 CGGCTGTTCTACGAGTGTGGATCACAATGAGTGTCTTATATATACAGGTACTGAACAG 2367  
Db 761 ArgLeuPheThrGluValLeuAspHisAsnGluCysLeuIleIleThrGlyThrGluGln 780  
QY 2368 TACAGTGAATACACAGGCTATGCTGAAACTTATCTGGGCGCCGAAAGCATGAAGATGG 2427  
Db 781 TyrSerGluTyrThrGlyTyrAlaGluThrTyrArgTyrAlaArgSerHisGluAspGly 800  
QY 2428 AGTGAAGAGCAGATTGGCAGCGCGCTGCGAGAGATCGTTGCCATTGACGCACTTCC 2487  
Db 801 SerGluLysAspAspTyrGlnArgCysThrGluIleValAlaIleAspAlaLeuHis 820  
QY 2488 TTGACAGCTACCTCGATCAGTTGTCTGCTGAGAGAGTGAAGCTGAGCTTAAACAGCT 2547  
Db 821 PheArgArgTyrLeuAspGlnPheValProGluLysValArgArgGluLeuAsnLysAla 840  
QY 2548 TACTGCGGATTCCTCGCTGAGAGTCTCTCTCAAAATCTTTCTGAGTGGCCACGCGA 2607  
Db 841 TyrCysGlyPheLeuArgProGlyValProSerGluAsnLeuSerAlaValAlaThrGly 860  
QY 2608 AACTGGGCTGTGTGCTTTTGGGGTGACGCTAGATTAAAGCCTTAATACAGATCTG 2667  
Db 861 AsnTyrGlyCysGlyAlaPheGlyAspAlaArgLeuLysAlaLeuIleGlnIleLeu 880  
QY 2668 GCACCTGCTGGCGCTGAACGTGACGTGTTTATTTTACCTTTGGGACTCAGATTGATG 2727  
Db 881 AlaAlaAlaAlaGluArgAspValValTyrPheThrPheGlyAspSerGluLeuMet 900  
QY 2728 AGACACATTTACAGCATGCACATTTCTTACCAGAGAGAGCTGGAGTTTGGAAAAGTG 2787  
Db 901 ArgAspIleTyrSerMetHisThrPheLeuThrGluArgLysLeuAspValGlyVal 920  
QY 2788 TACAAGTTATTGCTTAGATACATAAATGAAGAATGCAGAACTGTTCACCCCTGGACCA 2847  
Db 921 TyrLysLeuLeuLeuArgTyrTyrAsnGluGluCysArgAsnCysSerThrProGlyPro 940  
QY 2848 GACATCAAGCTTTTATCCATATACATCATCTGCTGCTGAGTCAAGTCAGACACCACTGAC 2907  
Db 941 AspIleLysLeuTyrProPheIleTyrHisAlaValGluSerSerAlaGluThrThrAsp 960  
QY 2908 ATGCAGACAGAGAAGCGAGGCACC 2931  
Db 961 MetProGlyGlnLysAlaGlyThr 968

## RESULT 3

US-09-511-507-6  
; Sequence 6, Application US/09511507  
; Patent No. 6395543  
; GENERAL INFORMATION:  
; APPLICANT: JACOBSON, Myron K.  
; APPLICANT: JACOBSON, Elaine L.  
; APPLICANT: AME, Jean-Christophe  
; APPLICANT: LIN, Winston  
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE (PARG) EN  
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV  
; TITLE OF INVENTION: THEREWITH

```

; FILE REFERENCE: NIAD 201
; CURRENT APPLICATION NUMBER: US/09/511,507
; CURRENT FILING DATE: 2000-02-23
; PRIOR APPLICATION NUMBER: 09/302,812
; PRIOR FILING DATE: 1999-04-30
; NUMBER OF SEQ ID NOS: 38
; SEQ ID NO 6
; LENGTH: 968
; TYPE: PRT
; ORGANISM: Mus musculus
; FEATURE:
; US-09-511-507-6

Alignment Scores:
  Pred. No.: 0          Length: 968
  Score: 5164.00        Matches: 968
  Percent Similarity: 100.00%  Conservative: 0
  Best Local Similarity: 100.00%  Mismatches: 0
  Query Match: 76.24%          Indels: 0
  DB: 4                     Gaps: 0

US-09-302-812-5 (1-3814) x US-09-511-507-6 (1-968)

QY 28 ATGAGTGGGGCCCGCTGGAGCCCTGCACAAAGCCGCTGGGGCCGCGTGAAC 87
DB 1 MetSerAlaGlyProGlyTrpGluProCysThrLysAlaArgTrpGlyAlaAlaGlyThr 20
QY 88 TCTGCGCCGACTCGCTCGGACTCCCGAGCTCCCTGGCAGGAGGCGGTCTTCGAC 147
DB 21 SerAlaProThrAlaSerAspSerArgSerPheProGlyArgGlnArgValLeuAsp 40
QY 148 CCCAAGGAGCGTCCCGTCCAGTTCAGGGTCCCTCCCTCGCCAGCCTGCGTCTCGGG 207
DB 41 ProLysAspAlaProValGlnPheArgValProProSerSerProAlaCysValSerGly 60
QY 208 CGGGGGGAGCCGACAGAGGACGACCTCGTGTGTTGTTTCAACAAAGACTATTCT 267
DB 61 ArgAlaGlyProHisArgGlyAsnAlaThrSerPheValPheLysGlnLysThrIleThr 80
QY 268 ACTTGGATGATTAAGGACCCAGACAGCTGAATCAGAAAGTAAAGAAACACAACT 327
DB 81 ThrTrpMetAspThrLysGlyProLysThrAlaGluSerGluSerLysGlnAsnAsn 100
QY 328 ACAAGATTGACTCCATGATGAGTTCGTGCGAAGATAACTTTTACCACATAGGTG 387
DB 101 ThrArgLeuAspSerMetSerValGlnLysAspAsnPheTyrProHisLysVal 120
QY 388 GAAAAATTGGAAATGTTCTCAGCTAAATCTGTATAATCACCACAGAAAGATTCA 447
DB 121 GluLysLeuGluAsnValProGlnLeuAsnLeuAspLysSerProThrGluLysSer 140
QY 448 CAGTATTTGAACCAACAGCAGACTGCGAGTGTGCAAGTGCAGAAATGAAGGAGCAT 507
DB 141 GlnTyrLeuAsnGlnGlnThrAlaSerValCysLysLysTrpGlnAsnGluLysHis 160
QY 508 CGAGAACAGCTTTTGGCAGTGAAGTCCCGGGGACTCCGCTACCAAGCAGCTTAGT 567
DB 161 AlaGluGlnLeuLeuAlaSerGluProProAlaGlyThrProLeuProLysGlnLeuSer 180
QY 568 ATGTCTAATGTTGGTCACTCCACACTGATGACCAAGTGCAGTGACACAGATCATGAAGA 627
DB 181 AsnAlaAsnIleGlyGlnSerProHisThrAspAspHisSerAspThrAspHisGluGlu 200
QY 628 GACAGACAAATCAGCAGTTCTTACACCTATAAACTTGAACATACAAAGCCACAGTA 687
DB 201 AspArgAspAsnGlnGlnPheLeuThrProIleLysLeuAlaAsnThrLysProThrVal 220
QY 688 GGAGATGGGCGGACGAGCAACTGTAGTGCAGTGGATCTCGCCAGTCTGTGAAGAC 747
DB 221 GlyAspGlyGlnAlaArgSerAsnCysLysCysSerGlySerArgGlnSerValLysAsp 240
QY 748 TGTACAGGCTGTCAACAGAGAGGAGTGGATGCTACAGAGAGTCTTTGTGACATGTT 807

```

```

DB 241 CysThrGlyCysGlnGlnGluValAspValLeuProGluSerProLeuSerAspVal 260
QY 808 GGTGCGGAGCAGCATTTGGAATCGACCCAAAATATGACAAACAATTTGACTGCGAAGAAC 867
DB 261 GlyAlaGluAspIleGlyThrGlyProLysAsnAspAsnLysLeuThrGlyGlnGluSer 280
QY 868 AGCTAGGTGATTGCGCTCCATTTGAGAAAGAAAGTGAGCCTGAGTCACCAATGGATGTA 927
DB 281 SerLeuGlyAspSerProPheGluLysGluSerGluProGluSerProMetAspVal 300
QY 928 GACAACTCGAGAAACAGTTGTCAAGATTCAGAGCAGATGCAAGAAACAAGTCAGTCTTT 987
DB 301 AspAsnSerArgAsnSerCysGlnAspSerGluAlaAspGluGluThrSerProValPhe 320
QY 988 GATGACGAGATGATCGTTCTCCCAACAGCAGCAATAAATCTTCAAGTTGCCAAGCAAGA 1047
DB 321 AspGlnGlnAspAspArgSerGlnThrAlaAsnLysLeuSerCysGlnAlaArg 340
QY 1048 GAAGCTGATGGCGATCTTAGGAAACCGTATTTGACTTAAGCGAAGTGAAGTTAGATTGCAT 1107
DB 341 GluAlaAspGlyAspLeuArgLysArgTyrLeuThrLysGlySerGluValArgLeuHis 360
QY 1108 TTCCAATTTGAAGGAAATAATATCTGGGACAGTGACTTAATAATGCCAAGCAATCTGGA 1167
DB 361 PheGlnPheGluGlyGluAsnAsnAlaGlyThrSerAspLeuAsnAlaLysProSerGly 380
QY 1168 AACTCTTCTAGCTTAATGTCAGAGTGTAGAGTCCCAAGCAGCATGCAAAAAGGGATTCT 1227
DB 381 AsnSerSerSerLeuAsnValGluCysArgSerSerLysGlnHisGlyLysArgAspSer 400
QY 1228 AAAATTCACAGATCATTTTCATGAGAAATTTCCAAAGTCAGAGCAGAGAAAGAAACAACTGT 1287
DB 401 LysIleThrAspHisPheMetArgIleSerLysSerGluAspArgLysGluGlnCys 420
QY 1288 GAAGTCAGATCAAGAAACAGAAAGGAGATCCAAATATACATCCACCTAACCTCCCT 1347
DB 421 GluValArgHisGlnArgThrGluArgLysIleProLysTyrIleProAsnLeuPro 440
QY 1348 CCAGAGAGAAAGTGTGGAACTCTATTGAGGAAATCAGAAAAATGCTCGGTGGG 1407
DB 441 ProGluLysLysTrpLeuGlyThrProIleGluGluMetArgLysMetProArgCysGly 460
QY 1408 ATCCATTTGCCCTTCCTTAAGACCATCTCGAAGTCACAGTGAAGTGTTCGGGTAGACCTT 1467
DB 461 IleHisLeuProSerLeuArgProSerAlaSerHisThrValThrValArgValAspLeu 480
QY 1468 CTGACAGCAGGAGAGTTCGGAACCTTTCCAAACATTAACAAGATTGTTGGGATAC 1527
DB 481 LeuArgAlaGlyGluValProLysProPheProThrHisTyrLysAspLeuTrpAspAsn 500
QY 1528 AAACATGTGAAAAATGCCCTTTGCGAAACAAAACTTGTACCTGTGGAAGATGAGATGCT 1587
DB 501 LysHisValLysMetProCysSerGluGlnAsnLeuTyrProValGluAspGluAsnGly 520
QY 1588 GAGCAACTCGAGGAGTAGTGGAGCTCATTCAGATGCACTTCTCAACAAATTCACA 1647
DB 521 GluArgThrAlaGlySerArgTrpGluLeuIleGlnThrAlaLeuLeuAsnLysPheThr 540
QY 1648 CGACCCCGAAGTTCGAGATGCGATTCGAAATCAATGTCGATATTCCTAAGAAATGG 1707
DB 541 ArgProGlnAsnLeuLysAspAlaIleLeuLysTyrAsnValAlaTyrSerLysLysTrp 560
QY 1708 GACTTTACAGCTTTGTTGATTCTTGGGATTAAGGTACTTGAAGAGCAGAGCCCAACAT 1767
DB 561 AspPheThrAlaLeuValAspPheTrpAspLysValLeuGluGluAlaGluAlaGlnHis 580
QY 1768 TTATATCAGTCCATTTTACCTGACATGCTGTAATTCGACTCTGTCTGCCAAATATTTGC 1827
DB 581 LeuTyrGlnSerIleLeuProAspMetValLysIleAlaLeuCysLeuProAsnIleCys 600
QY 1828 ACCGACCAATACACTCTCTGAAACAGAGATGAATCATTTCTGTCACGATGTACAGAA 1887
DB 601 ThrGlnProIleProLeuLeuLysGlnLysMetAsnHisSerValThrMetSerGlnGlu 620

```



Db 180 GluGlnPheSerAsnAlaAsnIleAspArgSerProGlnAsnAspAspHisSerAspThr 199  
 QY 616 GATCATGAAGAAGACAGAGCAATACAGCTTCTTACACCTATAAACTTGCAATACAA 675  
 Db 200 AspSerGluGluAsnArgAspAsnGlnGlnPheLeuThrThrValIysLeuAlaAsnAla 219  
 QY 676 AAGCCAACTAGGAGATGGCAGGCCAGAG-----AGCAACTGTAGTGCAGTGGA 726  
 Db 220 LysGlnThrThrGluAspGluHisAlaArgGluAlaLysSerHisGlnLysCysSerLys 239  
 QY 727 TCTCGCCAGCTCTGTGAAGACTGTACAGGCTGTCAACAGGAGGAGGTGGATGTGCTACCA 786  
 Db 240 SerCysHisProGlyGluAspCysAlaSerCysGlnGlnAspGluIleAspValPro 259  
 QY 787 GAGAGTCTTTGTCAGATGTTGTCGCGAGGACATTGGAACTGACACCAAAAAATGCAAC 846  
 Db 260 LysSerProLeuSerAspValGlySerGluAspValGlyThrGlySerLysAsnAspAsn 279  
 QY 847 AAATTGACTGGACAGAAAGCAGCCTAGGTGATTGCTCCTCATTTGAGAAAGAGTGAG 906  
 Db 280 LysLeuIleArgGlnGlnSerCysLeuGlyAsnSerProPheGlnLysGluSerGlu 299  
 QY 907 CCTGAGTCAACCAATGATGTAGACAACTCGAGAAACAGATTGTCAAGATTTCAGAGCAGAT 966  
 Db 300 ProGluSerProMetAspValAspAsnSerLysAsnSerCysGlnAspSerGluAlaAsp 319  
 QY 967 GAAGAAACAGTCCAGTCTTTGATGAGCAAGATGAT---CGTTCTCCCAACAGCAAAAT 1023  
 Db 320 GluGluThrSerProGlyPheAspGluGlnGluAspGlySerSerSerGlnThrAlaAsn 339  
 QY 1024 AAATTTCAAGTTGTCACAGCAGAGCTGTAGCGCATCTTAGGAACGGTATTGACT 1083  
 Db 340 LysProSerArgPheGlnAlaArgAspAlaAspIleGluPheArgLysArgTyrSerThr 359  
 QY 1084 AAGGGAGTGAAGTTAGATTGCAATTTCCAAATTTGAA---CGAGAAATATATGCTGGACC 1140  
 Db 360 LysGlyGlyGluValArgLeuHisPheGlnPheGluGlyGlyGluSerArgThrGlyMet 379  
 QY 1141 AGTGACTTAATGCCAAGCCATCTGAAACTCTTCTAGCTTAACTAGAGTGTAGAAGT 1200  
 Db 380 AsnAspLeuAsnAlaLysLeuProGlyAsnIleSerSerLeuAsnValGluCysArgAsn 399  
 QY 1201 TCCACGACGATGGAAAAAGGGATTTCAAATACAGATCATTTTCATGAGATTTCCAAG 1260  
 Db 400 SerLysGlnHisGlyLysSerLysIleThrAspHisLeuMetArgLeuProLys 419  
 QY 1261 TCAGAGGACAGAAAGAAACAATCTGAAGTCAGACATCAAGAAACAGAAAGAAAGATT 1320  
 Db 420 AlaGluAspArgArgLysGluGlnTrpGluThrLysHisGlnArgThrGluArgLysIle 439  
 QY 1321 CCABAATACATCCCACTAACCTCCCTCCAGAGAAAGTGGTGGAACTCCTATTGAG 1380  
 Db 440 ProLysTyrValProProHisLeuSerProAspLysLysThrPleuGlyThrProIleGlu 459  
 QY 1381 GAAATGAGAAAAATGCTCGGTGTGGATCCATTTGCTTCTTAAGACCATCTGCAAGT 1440  
 Db 460 GluMetArgArgMetProArgCysGlyLysLeuArgLeuProLeuLeuArgProSerAlaAsn 479  
 QY 1441 CACACAGTCACTGTTCCGGTAGACTCTCTCAGAGCAGGAGAGGTCCGAAACCTTTTCCA 1500  
 Db 480 HisThrValThrIleArgValAspLeuLeuArgAlaGlyGluValProLysProPhePro 499  
 QY 1501 ACACATTACAAAGATTGTGGGNTAACAAACATGTGAAATGCTTGTTCGGAACAAAC 1560  
 Db 500 ThrHisTyrLysAspLeuTrpAspAsnLysHisValLysMetProCysSerGluGlnAsn 519  
 QY 1561 TTGTACCTGTGGAAGATGTGAGCGCAACTGCGAGGAGTGAAGTGGAGCTCAT 1620  
 Db 520 LeuTyrProValGluAspGluAsnGlyGluArgThrAlaGlySerArgTrpGluLeuIle 539  
 QY 1621 CAGACTGCATCTCTCAACAAATTTCAACAGACCCCAAGACTTGAAGGATCGATTCTGAAA 1680  
 Db 540 GlnThrAlaLeuLeuAsnLysPheThrArgProGlnAsnLeuLysAspAlaIleLeuLys 559

QY 1681 TACATGTGCGCATATTCTTAAGAAATGGGACTTTACAGCTTTGGTTGATTCTGGGATAAG 1740  
 Db 560 TyrAsnValAlaTyrSerLysLysTrpAspPheThrAlaLeuIleAspPheTrpAspLys 579  
 QY 1741 GTACTTTGAAGAAGCAGAGCCCAACATTTATATACGTCCATTTTACCTGACATGGTGAAA 1800  
 Db 580 ValLeuGluGluAlaGluAlaGlnHisLeuTyrGlnSerIleLeuProAspMetValLys 599  
 QY 1801 ATTGCATCTGTCTGCCAAATATTTCACCCACCAATACCACTCTCTGAAACAGAGATG 1860  
 Db 600 IleAlaLeuCysLeuProAsnIleCysThrGlnProIleProLeuLeuLysGlnLysMet 619  
 QY 1861 AATCATCTGTCTCAGATGTCAAGAAACAGATCGCCAGCTTTTGTAGTAAATGTTCTTC 1920  
 Db 620 AsnHisSerIleThrMetSerGlnGluGlnIleAlaSerLeuLeuAlaAsnAlaPhePhe 639  
 QY 1921 TGCACATTTCCCGACCGAATGCCAAGATGCAATCGGAGTATTCTAGTTACCCAGACATT 1980  
 Db 640 CysThrPheProArgAsnAlaLysMetLysSerGluTyrSerSerTyrProAspIle 659  
 QY 1981 AACTTCAATCGTGTGTTGAAGGACCTTCATCAAGGAAACCCAGAAAACTGAAAAACATC 2040  
 Db 660 AsnPheAsnArgLeuPheGluGlyArgSerSerArgLysProGluLysLeuLysThrLeu 679  
 QY 2041 TTCTGTACTTTTCGAAGAGTCACAGAAAAAAACCTACAGGATTGTGTGACATTACAAGA 2100  
 Db 680 PheCysTyrPheArgArgValThrGluLysLysProThrGlyLeuValThrPheThrArg 699  
 QY 2101 CAGAGCTTTGAGATTTCCAGATCGGAAAGGTGTGAAGCGCTCTGACAGCTTACAC 2160  
 Db 700 GlnSerLeuGluAspPheProGluTrpGluArgCysGluLysProLeuThrArgLeuHis 719  
 QY 2161 GTCACCTTACAGAGGTACCATAGAGCAACGGCCAGGAGCATGTACAGTGGATTGTCGA 2220  
 Db 720 ValThrTyrGluGlyThrIleGluGluAsnGlyGlnGlyMetLeuGlnValAspPheAla 739  
 QY 2221 AATCGTTTGTGGAGTGGTGTGACTGTGTGGGGACTTGTACAAAGAAATCAGATT 2280  
 Db 740 AsnArgPheValGlyGlyGlyValThrSerAlaGlyLeuValGlnGluGluIleArgPhe 759  
 QY 2281 TTAATCAATCTGAATTGATTGTTTACCGCTGTTCACGTGAGTGTGATGATCAATGAG 2340  
 Db 760 LeuIleAsnProGluLeuIleSerArgLeuPheThrGluValLeuAspHisAsnGlu 779  
 QY 2341 TGTCTTATTATCACAGGTACTGAACAGTACAGTGAATACACAGGCTATGCTGAACTTAT 2400  
 Db 780 CysLeuIleIleThrGlyThrGluGlnTyrSerGluTyrThrGlyTyrAlaGluThrTyr 799  
 QY 2401 CGTTGGCCCGAGCCATGAAGTGGAGTGAAGAGGAGTGGCAGCGCGCTGCACG 2460  
 Db 800 ArgTrpSerArgSerHisGluAspGlySerGluArgAspAspCysGluArgCysThr 819  
 QY 2461 GAGATCGTTGCCATTGACGCACTTCATTCAGACGCTACTCGATCAGTTGTGCTGAG 2520  
 Db 820 GluIleValAlaIleAspAlaLeuHisPheArgArgTyrLeuAspGlnPheValProGlu 839  
 QY 2521 AAAGTGAGAGTGTGAGCTTAACAGGCTTACTCGGATTCCTCGCTCGCTGAGTTCCTT 2580  
 Db 840 LysMetArgArgGluLeuAsnLysAlaTyrCysGlyPheLeuArgProGlyValSerSer 859  
 QY 2581 GAAATCTTCTGAGTGGCCAGGAACTGGGGCTGGTGGCTTGGGGTGGAGCT 2640  
 Db 860 GluAsnLeuSerAlaValAlaThrGlyAsnTrpGlyCysGlyAlaPheGlyGlyAspAla 879  
 QY 2641 AGATTAAAGCCCTTAATACAGATCTGCGAGCTGTGCGGCTGGAACGTGACGTGGTTTAT 2700  
 Db 880 ArgLeuLysAlaLeuIleGlnLeuAlaAlaAlaGluArgAspValValTyr 899  
 QY 2701 TTCACCTTTGGGACTCAGAGTTGATGAGACATTTACAGCATGCACATTTCTCTTACC 2760  
 Db 900 PheThrPheGlyAspSerGluLeuMetArgAspIleTyrSerMetHisIlePheLeuThr 919







QY 1501 ACACATTACAAAGATTGTGGGATAACAAACATGCTGAAATGCTGTGTGGGAAACAAAC 1560  
Db 500 ThrHisTyrLysAspLeuTyrPaspAsnLysHisValLysMetProCysSerGluGlnAsn 519  
QY 1561 TTGTACCTCTGGAGAGTGAAGTGGTGGAGCAACTGCGAGGAGTGGTGGAGCTCAAT 1620  
Db 520 LeuTyrProValGluAspGluAsnGlyGluArgThrAlaGlySerArgTyrGluLeuLeu 539  
QY 1621 CAGACTGCCTCTCAACAAATTCACAGCAGCCAGCACTTGAAGGATGCAATTCGAAA 1680  
Db 540 GlnThrAlaLeuLeuAsnLysPheThrArgProGlnAsnLeuLysAspAlaLeuLys 559  
QY 1681 TACAATGTGGCATATCTTAAGAAATGGGACTTTACAGCTTTGGTGTGATTTCTGGGATAAG 1740  
Db 560 TyrAsnValAlaTyrSerLysLysTyrAspPheThrAlaLeuLeuAspPheTyrAspLys 579  
QY 1741 GTACTTGAAGAGCAGAGGCGCAATTTATATCAGTCCATTTTACCTGATGCTGAAA 1800  
Db 580 ValLeuGluGluAlaGluAlaGlnHisLeuTyrGlnSerIleLeuProAspMetValLys 599  
QY 1801 ATTGCACCTGTCTGCCAAATATTGACCCAGCCAGCAATACCACTCTGAAACAGAAATG 1860  
Db 600 IleAlaLeuCysLeuProAsnIleCysThrGlnProIleProLeuLeuLysGlnLysMet 619  
QY 1861 AATCAATCTCTGCAGATGTCACAGAAACAGATCCCACTGCTTTTAGCTAATGCTTCTTC 1920  
Db 620 AsnHisSerIleThrMetSerGlnGluGlnIleAlaSerLeuLeuAlaAsnAlaPhePhe 639  
QY 1921 TGACATTTCCCGACGGAATGCCAAGATGAATCGGAGTATTCAGTACCCAGACATT 1980  
Db 640 CysThrPheProArgArgAsnAlaLysMetLysSerGluTyrSerTyrProAspIle 659  
QY 1981 AACTCAATCGTGTGTGAGAGCGTTCATCAAGAAACACAGAAACCTGAAACCACTC 2040  
Db 660 AsnPheAsnArgLeuPheGluGlyArgSerSerArgLysProGluLysLeuLysThrLeu 679  
QY 2041 TTCTGCTACTTTTCCAGAGTCACAGAAACAACTACAGATTTGGTGAATTTACAAGA 2100  
Db 680 PheCysTyrPheArgArgValThrGluLysLysProThrGlyLeuValThrPheThrArg 699  
QY 2101 CAGAGTCTTGAAGATTTTCCAGAAATGGGAAGGTGTGAAAGCTCTGACACGCTTACAC 2160  
Db 700 GlnSerLeuGluAspPheProGluTyrGluArgCysGluLysProLeuThrArgLeuHis 719  
QY 2161 GTCACTTTACAGAGGTACCATAGAGGCAACGCGCGGCGATGTACAGGTGGATTTGCA 2220  
Db 720 ValThrTyrGluGlyThrIleGluGlnAsnGlyGlnGlyMetLeuGlnValAspPheAla 739  
QY 2221 AATCGTTTTTGTGGAGTGTGTGACTGTGCGGACTTGTACAAGAAAGAAATCAGATTT 2280  
Db 740 AsnArgPheValGlyGlyGlyValThrSerAlaGlyLeuValGlnGluGluIleArgPhe 759  
QY 2281 TTAATCAATCTGAATGTATGTTTACGCTGTTTACAGTGTGTGATCAATCAATAGAG 2340  
Db 760 LeuIleAsnProGluLeuIleIleSerArgLeuPheThrGluValLeuAspHisAsnGlu 779  
QY 2341 TGTCTTATTATACAGTACTACAGTACAGTGAATACACAGCTATGCTGAACTTAT 2400  
Db 780 CysLeuIleThrGlyThrGluGlnTyrSerGluTyrThrGlyThrAlaGluThrTyr 799  
QY 2401 CGTTGGCGCCGAAGCCATGAAGTGGAGTGAAGAGCAGCATTTGGCAGCGGCGCTGCAGC 2460  
Db 800 ArgTyrSerArgSerHisGluAspGlySerGluArgAspAspCysGluArgArgCysThr 819  
QY 2461 GAGATGTTGCCATTCAGCGCACTTCACTTACAGAGCTACCTCATCATGTTTGGCTGAG 2520  
Db 820 GluIleValAlaIleAspAlaLeuHisPheArgArgTyrLeuAspGlnPheValProGlu 839  
QY 2521 AAAGTGAAGCGTCAAGAGCTTACTCGGGATTCCTCGCTCGTGGAGTTCCTTCT 2580  
Db 840 LysMetArgArgGluLeuAsnLysAlaTyrCysGlyPheLeuArgProGlyValSerSer 859

QY 2581 GAAATCTTTCTGCAGTGGCCACGGGAAACTGGGGCTGTGGTCTGCTTTGGGGGTGACGCT 2640  
Db 860 GluAsnLeuSerAlaValAlaThrGlyAsnTyrGlyCysGlyAlaPheGlyGlyAspAla 879  
QY 2641 AGATTAAAGCCCTTAATACAGATCTCTGCGAGCTCTGCGGCTGAACCTGACGCTGTTAT 2700  
Db 880 ArgLeuLysAlaLeuIleGlnIleLeuAlaAlaAlaGluArgAspValValTyr 899  
QY 2701 TTCACCTTTGGGAGTCTCAGAGTGTGATGAGAGACATTTACAGCATGCACTTCTCTTACC 2760  
Db 900 PheThrPheGlyAspSerGluMetArgAspIleTyrSerMetHisIlePheLeuThr 919  
QY 2761 GAGAGGAAGCTGCAATGTTGAAAAAGTGTACAAGTATTGCTTGTAGTACTACTACAATGAAGAA 2820  
Db 920 GluArgLysLeuThrValGlyAspValTyrLysLeuLeuLeuArgTyrTyrAsnGluGlu 939  
QY 2821 TGCAGAAATCTGTTCCACCCCTGGACACAGACATCAAGCTTTTATCCATTCATATACCATGCT 2880  
Db 940 CysArgAsnCysSerThrProGlyProAspIleLysLeuTyrProPheIleTyrHisAla 959  
QY 2881 GTTGAGTCAAGTGCAGAGCACCTGACATCCAGGACAGAGGCGGACCC 2931  
Db 960 ValGluSerCysAlaGluThrAlaAspHisSerGlyGlnArgThrGlyThr 976

RESULT 6  
US-09-511-507-4  
; Sequence 4, Application US/09511507  
; Patent No. 6395543  
; GENERAL INFORMATION:  
; APPLICANT: JACOBSON, Myron K.  
; APPLICANT: JACOBSON, Elaine L.  
; APPLICANT: AME, Jean-Christophe  
; APPLICANT: LIN, Winston  
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOPOLYMERASE (PARG) EN  
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIVE  
; FILE REFERENCE: NIAD 201  
; CURRENT APPLICATION NUMBER: US/09/511,507  
; PRIOR FILING DATE: 2000-02-23  
; PRIOR FILING DATE: 09/302,812  
; NUMBER OF SEQ ID NOS: 38  
; SEQ ID NO 4  
; LENGTH: 976  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
US-09-511-507-4

Alignment Scores:  
Pred. No.: 0 Length: 976  
Score: 4395.00 Matches: 837  
Percent Similarity: 90.99% Conservative: 52  
Best Local Similarity: 85.67% Mismatches: 78  
Query Match: 64.89% Indels: 10  
DB: 4 Gaps: 6

US-09-302-812-5 (1-3814) x US-09-511-507-4 (1-976)

QY 28 ATGAGTGGCGGCGCCCGCTGGAGCGCTGCAGGAAAGCG---CGCTGGGCGCGCGCTGGA 84  
Db 1 MetAsnAlaGlyProGlyCysGluProCysThrLysAlaThrArgTyrGlyAlaAla--- 19  
QY 85 ACTTCTGCGCGGAGTCCCTCGGAGTCCCGAGCTCCCTGGCAGGAGGCGGTCTCTC 144  
Db 20 ThrThrSerProAlaAlaSerAspAlaArgSerPheProSerArgGlnArgValLeu 39  
QY 145 GACCCCAAGAGCGCTCCCGTCCAGTTCAGGGTCCCTCCCTCCGCGAGCTCGCTCG 204  
Db 40 AppProLysAspAlaHisValGlnPheArgValProProSerSerProAlaCysValPro 59  
QY 205 GGGCGGCGGAGCCCGCACAGAGGCAACGCCACCTCGTTGTTTTCACAAAGACTATT 264  
Db 205 GGGCGGCGGAGCCCGCACAGAGGCAACGCCACCTCGTTGTTTTCACAAAGACTATT 264

Db 60 GlyGlnAlaGlyGlnHisArgGlySerAlaThrSerLeuValPheLysGlnLysThrIle 79  
QY 265 ACTACTGGATGATCTAAGACCCAGACAGCTGAATCAGAA-----AGTAAA 315  
Db 80 ThrSerTrpMetAspThrLysGlyIleLysThrAlaGluSerGluSerLeuAspSerLys 99  
QY 316 GAAAAACAATAACAAGATTGACTCCATGATCAGTCTGTGCAGAGAAAGATACCTTTTAC 375  
Db 100 GluAsnAsnThrArgIleGluSerMetMetSerValGlnLysAspAsnPheThr 119  
QY 376 CCACATAAGTGGAAAAATGGAAAAATGTTCTCAGCTAAATCTTGATAAATCACCCACA 435  
Db 120 GlnHisAsnValGlnLysLeuValAsnValSerGlnLeuSerLeuAspLysSerLeuThr 139  
QY 436 GAAAGAGTTCAGATATTGAAACCAACAGCAGACTGGAGTGTGTGCGAGTGGCAGAT 495  
Db 140 GluLysSerThrGlnTrpLysLeuAsnGlnHisGlnThrAlaAlaMetCysLysTrpGlnAsn 159  
QY 496 GAAGGGAAGCATGCAGAACAGCTTTTGGCAAGTGAGCCTCCCGCGGGGACTCCGCTACCA 555  
Db 160 GluGlyLysHisThrGluGlnLeuLeuGluSerGluProGlnThrValThrLeuValPro 179  
QY 556 AAGCAGCTTAGTAATGCTTAACATGTGTAGTCACTCCACACACTGATGACACAGTGACACA 615  
Db 180 GlnGlnPheSerAsnAlaAsnIleAspArgSerProGlnAsnAspAspHisSerAspThr 199  
QY 616 GATCATGAAGACAGACAGACATCAGCATCTTCTTACACCTATATAAACTTGCATAATACA 675  
Db 200 AspSerGluGluAsnArgAsnGlnGlnPheLeuThrValLysLeuAlaAsnAla 219  
QY 676 AAGCCAACATAGGAGATGGCAGGCGAGA-----AGCAACTGTAAAGTGCAAGTGA 726  
Db 220 LysGlnThrThrGluAspGluHisAlaArgGluAlaLysSerHisGlnLysCysSerLys 239  
QY 727 TCTCGCCAGTCTGTGAAGACTGTGACGCTGTCAACAGGAGAGTGGATGTGCTACCA 786  
Db 240 SerCysHisProGlyGluAspCysAlaSerCysGlnGlnAspGluIleAspValPro 259  
QY 787 GAGACTCTTTGTCAGATTTGGTCCGAGGACATTTGAACTGGACCAAAAATGACAAAC 846  
Db 260 LysSerProLysSerAspValGlySerGluAspValGlyThrGlySerLysAsnAspAsn 279  
QY 847 AAATGACTGCAACAAGAACAGCTAGGTGATTCGCTCCATTTGAGAAAACAAAGTGA 906  
Db 280 LysLeuIleArgGlnGluSerCysLeuGlyAsnSerProPheGlnLysGluSerGlu 299  
QY 907 CTGTAGTCACCAATGGATGTAGCACTCGAAGAACAGTGTCAAGATTCAGAACAGAT 966  
Db 300 ProGluSerProMetAspValAspAsnSerLysAsnSerCysGlnAspSerGluAlaAsp 319  
QY 967 GAAGAAACAAGTCCAGTCTTTGATGACGCAAGATGAT---CGTCTCTCCCAACAGCAAT 1023  
Db 320 GluGluThrSerProGlyPheAspGluGlnGluAspGlySerSerGlnThrAlaAsn 339  
QY 1024 AAACCTTCAAGTTCGAAGAGAGAGCTGATGGCATCTTAGGAAACGGTATTGTACT 1083  
Db 340 LysProSerArgPheGlnAlaArgAspAlaAspIleGluPheArgLysArgTyrSerThr 359  
QY 1084 AAGGAAGTGAAGTTAGATTTGATTTCCAAATTTGAA---GAGAAAAATAAGTCTGGGACC 1140  
Db 360 LysGlyGlyGluValArgLeuHisPheGlnPheGluGlyGlyGluSerArgThrGlyMet 379  
QY 1141 AGTGACTTAAATGCCAAGCCATCTCGAAACTCTTACGCTTAAGTAGATGTAGAAGT 1200  
Db 380 AsnAspLeuAsnAlaLysLeuProGlyAsnIleSerSerLeuAsnValGluCysArgAsn 399  
QY 1201 TCCAGACGACATGAAAAAGGATTTCTAAATTTACAGATCATTTTCATGAGAAATTTCCAA 1260  
Db 400 SerLysGlnHisGlyLysLysAspSerLysIleThrAspHisLeuMetArgLeuProLys 419  
QY 1261 TCAGAGGACAGAGAAAAGAACAAATGTGAAGTTCAGATCAATCAAGAACAGAGAGAT 1320  
Db 420 AlaGluAspArgArgLysGluGlnTrpGluThrLysHisGlnArgThrGluArgLysIle 439

QY 1321 CCATAATATCATCCACCTAACTCCCTCCAGAGAGAAAGTGGCTGGAACTCTCTATTGAG 1380  
Db 440 ProLysTrpValProProHisLeuSerProAspLysLysTrpLeuGlyThrProIleGlu 459  
QY 1381 GAAATGAGAAAAATGCCCTGGTGGATCCATTTCCCTTCTTAAGACCATCTGCAAGT 1440  
Db 460 GluMetArgMetProArgCysGlyIleArgLeuProLeuLeuArgProSerAlaAsn 479  
QY 1441 CACACAGTGAATCTCGGGTAGACCTTCTGAGAGCAGAGAGGTTCCGAAACCTTTTCCA 1500  
Db 480 HisThrValThrIleArgValAspLeuLeuArgAlaGlyGluValProLysProPhePro 499  
QY 1501 ACACATTACAAGATTTGTGGATACAACAATGTGAAAATCCCTTTGCGAAACAAAAC 1560  
Db 500 ThrHisTrpLysAspLeuTrpAspAsnLysHisValLysMetProCysSerGluGlnAsn 519  
QY 1561 TTGTACCTGTGGAAGATGAGAACTGTGAGCAACTGCAGGAGTAGGTGGGAGCTCAT 1620  
Db 520 LeuTrpProValGluAspGluAsnGlyGluArgThrAlaGlySerArgTrpGluLeuIle 539  
QY 1621 CAGACTGCATCTCTCAACAATTCACAGACCCCAAGAACTTGAAGGATGCGATTCTGAAA 1680  
Db 540 GlnThrAlaLeuLeuAsnLysPheThrArgProGlnAsnLeuLysAspAlaIleLeuLys 559  
QY 1681 TACATGTGGCATATCTCAAGAAATGGGACTTTACAGCTTTGGTGTGATTTCTGGGATAG 1740  
Db 560 TyrAsnValAlaTrpSerLysLysTrpAspPheThrAlaLeuIleAspPheTrpAspLys 579  
QY 1741 GTACTTGAAG 1800  
Db 580 ValLeuGluGluAlaGluAlaGlnHisLeuTrpGlnSerIleLeuProAspMetValLys 599  
QY 1801 ATTGCACCTGTGTGCTCCAAATATTGTCACCCAGCCCAATACCACTCTCTGAAAACAGAGATG 1860  
Db 600 IleAlaLeuCysLeuProAsnIleCysThrGlnProIleProLeuLeuLysGlnLysMet 619  
QY 1861 AATCATTTCTGCAGATGTCTCAGAAACAGATCGCCAGCTCTTTTGTAGTAAATGCTTTCTTC 1920  
Db 620 AsnHisSerIleThrMetSerGlnGluGlnIleAlaSerLeuLeuAlaAsnAlaPhePhe 639  
QY 1921 TGACATTTTCCCGACGGAATGCCAAGATGAAATCGGAGTATCTAGTTACCCACACAT 1980  
Db 640 CysThrPheProArgArgAsnAlaLysMetLysSerGluTrpSerSerTrpProAspIle 659  
QY 1981 AACTCAATCGGTTGTTTGAAGAGCTTCATCAAGGAAACCCAGAAAACTGAAAAACATC 2040  
Db 660 AsnPheAsnArgLeuPheGluGlyArgSerSerArgLysProGluLysLeuLysThrLeu 679  
QY 2041 TTTCTGCTACTTTCCAGAGTTCACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2100  
Db 680 PheCysTrpPheArgArgValThrGluLysLysProThrGlyLeuValThrPheThrArg 699  
QY 2101 CAGAGTCTTGAAGATTTTCCAGAAATGGAAAGTGTGAAAAGCTCTGACACGCTTACAC 2160  
Db 700 GlnSerLeuGluAspPheProGluTrpGluArgCysGluLysProLeuThrArgLeuHis 719  
QY 2161 GTCCTTACAGAGGTACCATAGAGGCAACGCCGAGCATGCTACAGGTGATTTGCA 2220  
Db 720 ValThrTrpGluGlyThrIleGluGluAsnGlyGlnGlyMetLeuGlnValAspPheAla 739  
QY 2221 AATCGTTTGTGGAGGTGTGCTGCTGGTGGGACTTGTTCAGAGAGAGAGAGAGAGAGAG 2280  
Db 740 AsnArgPheValGlyGlyValThrSerAlaGlyLeuValGlnGluGluIleArgPhe 759  
QY 2281 TTAATCAATCTGAATTTGTTTCCAGCTGTTCCTGAGGTGCTGGATCACAATGAG 2340  
Db 760 LeuIleAsnProGluLeuLeuIleSerArgLeuPheThrGluValLeuAspHisAsnGlu 779  
QY 2341 TGTCTTATTATCACAGGTACTGAACAGTACAGTGAATACACAGCTATGCTGAAACTTAT 2400  
Db 780 CysLeuIleIleThrGlyThrGluGlnTrpSerGluTrpThrGlyTyrAlaGluThrTyr 799



QY	2221	AATCGTTTCTGCGAGGTGTGTGACATGCTGCGGACCTTGCTACAAGAAGAAATCAGATTT	2288
DB	741	AsnArgPheValGlyGlyValThrSerAlaGlyLeuValGlnGluGluIleArgPhe	760
QY	2281	TTAATCAATCCTCAATTGATTTTCACGGCTCTTCACTGAGTGTGGATACAAATGAG	2340
DB	761	LeuIleAsnProGluLeuIleValSerArgLeuPheThrGluValLeuAspHisAsnGlu	780
QY	2341	TGCTTTATTATCACAGTACTGAACACAGTACAGTGAATACACAGGCTATGCTGAAACTTAT	2400
DB	781	CysLeuIleIleThrGlyThrGluGlnTyrSerGluTyrThrGlyTyrAlaGluThrTyr	800
QY	2401	CGTTGGCCCGAAGCCATGAAGATGGAGTGAAGAGCAAGATTGGCAGCGCGCTGCACG	2460
DB	801	ArgTrpAlaArgSerHisGluAspArgSerGluArgAspTrpGlnArgArgThrThr	820
QY	2461	GAGATCGTTGCCAATTCACGCACTTCCTACTCAGACGCTACTCCATCAGTTGTGCTCTGAG	2520
DB	821	GluIleValAlaIleAspAlaLeuHisPheArgArgTyrLeuAspGlnPheValProGlu	840
QY	2521	AAAGTGAGAGCTGAGCTTTAAACAGGCTTACTTGGGAAATTCCTCCGTCTGTGAGTCTTCT	2580
DB	841	LysIleArgArgGluLeuAsnLysAlaTyrCysGlyPheLeuArgProGlyValSerSer	860
QY	2581	GAATAATCTTCTCAGTGCACCGGAAACTGGGGCTGTGGTGCCTTTCGGGGTGAAGCT	2640
DB	861	GluAsnLeuSerAlaValAlaThrGlyValAsnTrpGlyCysGlyAlaPheGlyGlyAspAla	880
QY	2641	AGATTAAAGCCTTAATACAGATTCCTGGCAGCTGCTGCGGCTGAACGTCACGTGGTTTAT	2700
DB	881	ArgLeuLysAlaLeuIleGlnIleLeuAlaAlaValAlaGluArgAspValValTyr	900
QY	2701	TTCACTTTGGGACCTCAGAGTTGATGAGAGACATTTACAGCATGACACTTTCCTTACC	2760
DB	901	PheThrPheGlyAspSerGluLeuMetArgAspIleTyrSerMethHisThrPheLeuThr	920
QY	2761	GACAGCAAGCTGATCTTCGAAAAGTGACAACTTATTCCTTAGATACACTACAATGAAGA	2820
DB	921	GluArgLysLeuThrValGlyGluValTyrLysLeuLeuLeuArgTyrTyrAsnGluGlu	940
QY	2821	TGCAGAAACTGTTCACCCCTGCAGCACATCAAGCTTTATCCATTATATACCATGCT	2880
DB	941	CysArgAsnCysSerThrProGlyProAspIleLysLeuTyrProPheIleTyrHisAla	960
QY	2881	GTTGACTCAAGTCGACAGACCACTGCATGCCAGGACGAGGACGGC	2928
DB	961	ValGluSerCysThrGlnThrThrAsnGlnProGlyGlnArgThrGly	976
RESULT 8			
US-09-511-477-2			
; Sequence 2, Application US/09511477			
; Patent No. 6337202			
; GENERAL INFORMATION:			
; APPLICANT: JACOBSON, Myron K.			
; APPLICANT: JACOBSON, Elaine L.			
; APPLICANT: ABE, Jean-Christophe			
; APPLICANT: LIN, Winston			
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE (			
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUN			
; FILE REFERENCE: NIAD 201			
; CURRENT APPLICATION NUMBER: US/09/511,477			
; CURRENT FILING DATE: 2000-02-23			
; PRIOR APPLICATION NUMBER: 09/302,812			
; PRIOR FILING DATE: 1999-04-30			
; NUMBER OF SEQ ID NOS: 38			
; SEQ ID NO 2			
; LENGTH: 977			
; TYPE: PRT			
; ORGANISM: Bos taurus			
; FEATURE:			
US-09-511-477-2			

Alignment Scores:  
 Pred. No.: 0 Length: 977  
 Score: 4332.50 Matches: 820  
 Percent Similarity: 89.86% Conservative: 57  
 Best Local Similarity: 84.02% Mismatches: 90  
 Query Match: 63.97% Indels: 9  
 DB: 4 Gaps: 5

US-09-302-812-5 (1-3814) x US-09-511-477-2 (1-977)

```

QY 28 ATGAGTCGGGCCCCCGCTGGAGAGCCCTGCACGAAA---GCGGCTGGGGCGCGCTGGA 84
Db 1 MetSerAlaGlyProGlyCysGluProCysThrLysArgProArgTrpAspAlaAla 20

QY 85 ACTTCTGCGCCGACTCCCTCGGACTCCCGAGCTCCCTGGGAGGAGGCGGTCTTC 144
Db 21 ThrSerProAlaAlaSerAspAlaArgSerPheProGlyArgGlnArgValLeu 40

QY 145 GACCCCAAGAGCTCCCTCCAGTTCAGGGTCCCTCCCTCGCTCGCAGCTCGCTCG 204
Db 41 AspSerLysAspAlaProValGlnPheArgValProProSerSerSerGlyCysAlaLeu 60

QY 205 GGGCGGCGGACCGCACAGAGCAACGCCACTCGTCTTTGTTTCAAAACAAAGACTATT 264
Db 61 GlyArgAlaGlyGlnHisArgGlySerAlaThrSerLeuValPheLysGlnYsThrIle 80

QY 265 ACTACTTGATGGATGATAAAGGACCCCAAGACAGCTGATCAGAA-----AGTAAA 315
Db 81 ThrSerTrpMetAspThrLysGlyIleYsThrValGluSerGluSerLeuHisSerLys 100

QY 316 GAAACAAACAATACAGAAATGACTCCATGATGAGTTCGTGCAGAAAGATAACTTTTAC 375
Db 101 GluAsnAsnThrArgGluGluSerMetMetSerSerValGlnYsAspAsnPheTyr 120

QY 376 CCACATAGGTGGAAAAATTGGAATGTTCTCAGCTAAATCTTGATTAATACCCACA 435
Db 121 GlnHisAsnMetGluLysLeuGluAsnValSerGlnLeuGlyPheAspLysSerProVal 140

QY 436 GAAAGAGTTTCACAGTATTGTAACCAACAGCAGCTGGCGAGTGTGCAAGTGGCAGAT 495
Db 141 GluYsGlyThrGlnYrLeuYsGlnHisGlnThrAlaAlaMetCysLysTrpGlnAsn 160

QY 496 GAAAGGAAGCATGACAGACTTTTGGCAAGTGGACCTCCCGCGGACTCCGCTACCA 555
Db 161 GluGlyProHisSerGluArgLeuLeuGluSerGluProProAlaValThrLeuValPro 180

QY 556 AAGCAGCTTAGTAATGCTAACTAGTCCAGTCCACCCACACTGATGACACAGTACACA 615
Db 181 GluGlnPheSerAsnAlaAsnValAspGlnSerSerProLysAspAspHisSerAspThr 200

QY 616 GATCATGAGAGACAGACAGCAATCAGCAGTTCCTTACACCTATATAACTTGCATACA 675
Db 201 AsnSerGluGluSerArgAspAsnGlnGlnPheLeuThrHisValLysLeuAlaAsnAla 220

QY 676 AAGCAACAGTAGGAGATGGGCGAG-----GCCAAGCAACTGTAAAGTGCAGTGA 726
Db 221 LysGlnThrMetGluAspGluGlnGlyArgGluAlaArgSerHisGlnLysCysGlyLys 240

QY 727 TCTGCCAGTCTGTGAAGACTGTTCAGGCTCTCAACAGGAGAGGTGATGTGCTACCA 786
Db 241 AlaCysHisProAlaGluAlaCysAlaGlyCysGlnGlnGluThrAspValValSer 260

QY 787 GAGAGTCTTGTTCAGATGTTGTCGCGAGGACATTGGAACTGGACCCAAAATGACAC 846
Db 261 GluSerProLeuSerAspThrGlySerGluAspValGlyThrGlyLeuLysAsnAlaAsn 280

QY 847 AAATTGACTGGAAGAAGACGCTTAGTGTATCGCTTCCATTTTCAGAAAGAAAGTGA 906
Db 281 ArgLeuAsnArgGlnGluSerSerLeuGlyAsnSerProProPheGluLysGluSerGlu 300

QY 907 CTGTGTCACCAATGGATGTAGCAACTCGAGAAACAGTTGTCAAGATTCAAGACAGAT 966
Db 301 ProGluSerProMetAspValAspAsnSerLysAsnSerCysGlnAspSerGluAlaAsp 320

```

```

QY 967 GAAGAAACAAGTCCAGTCTTTCATGAGCAAGATGATCGTTCCTCC---CAACAGCAAT 1023
Db 321 GluGluThrSerProGlyPheAspGluGlnGluAspSerSerAlaGlnThrAlaAsn 340

QY 1024 AAATCTTCAAGTTCGCAAGCAAGAGAGCTGATGGCGATCTTAGGAAACGGTATTGACT 1083
Db 341 LysProSerArgPheGlnProArgGluAlaAspThrGluLeuArgLysArgSerAla 360

QY 1084 AAGGAGAGTGAAGTAGATTGATTCATTTCCAATTGAA---GGAGAAATATGCTGGAGC 1140
Db 361 LysGlyGlyGluIleArgLeuHisPheGlnPheGluGlyGlyGluSerArgAlaGlyMet 380

QY 1141 ACTGACTTAAATGCCAAGCCACTCTCGAACTCTTAGCCTTAACTAGTAGTGAAGT 1200
Db 381 AsnAspValAsnAlaLysArgProGlySerThrSerSerLeuAsnValGluCysArgAsn 400

QY 1201 TCCAAAGCAGCATGGAAAAAGGATTTCTAAATATACAGATCATTTTCATGAGATTTCCAAG 1260
Db 401 SerLysGlnHisGlyArgLysAspSerLysIleThrAspHisPheMetArgValProLys 420

QY 1261 TCAGAGCAGCAGAGAAAGAACAAATGTGAAGTCAGACATCAAAAGACAGAAAGAGATT 1320
Db 421 AlaGluAspLysArgLysGluGlnCysGluMetLysHisGlnArgThrGluArgLysIle 440

QY 1321 CCAAAATACATCCCACTAACCTCCCTCCAGAGAAAGAGTGGCTGGAACTCTATTGAG 1380
Db 441 ProLysTyrIleProProHisLeuSerProAspLysLysIleProGlyThrProIleGlu 460

QY 1381 GAAATGAGAAAAATCCCTCGGTGGGATCCATTTGTCCTTCTTAAGACATCTGCAAGT 1440
Db 461 GluMetArgArgMetProArgCysGlyIleArgLeuProProLeuArgProSerAlaAsn 480

QY 1441 CACAGTGCAGTGTTCGGGTAGACTCTCTGAGAGCAGAGAGGTTCGGAACCTTTTCCA 1500
Db 481 HisThrValThrIleArgValAspLeuLeuArgIleGlyGluValProLysProPhePro 500

QY 1501 ACACATTACAAAGATTGTGGGATTAACAACATGTGAAATGCCTTGTTCGGAACAAAC 1560
Db 501 ThrHisPheLysAspLeuTrpAspAsnLysHisValLysMetProCysSerGluGlnAsn 520

QY 1561 TTGTACCTCTGGAAAGATGAGAAATCGTCAGCAACTGCAGGAGTAGTGGAGGCTCAT 1620
Db 521 LeuTyrProValGluAspGluAsnGlyGluArgAlaAlaGlySerArgTrpGluLeuIle 540

QY 1621 CAGACTGCATCTTCAACAAATTCACAGACCCCAAGACTTCGAGAGTCCGATTCGAAA 1680
Db 541 GlnThrAlaLeuLeuAsnArgLeuThrArgProGlnAsnLeuLysAspAlaIleLeuLys 560

QY 1681 TACATGTGGCATATTTCAAGAAATGGGACTTTACAGCTTTTGGTGTGATTTCTGGGATAG 1740
Db 561 TyrAsnValAlaTyrSerLysLysTrpAspPheThrAlaLeuIleAspPheTrpAspLys 580

QY 1741 GTACTTGAAGAGCAGAGGCCCAACATTTATATCAGTCCATTTTACCTGCATGCTGAAA 1800
Db 581 ValLeuGluGluAlaGluAlaGlnHisLeuTyrGlnSerIleLeuProAspMetValLys 600

QY 1801 ATTGCATCTCTCTGCCAAATATTGGACCCAGCCCAATACCACTCTCGAAACAGAGATG 1860
Db 601 IleAlaLeuCysLeuProAsnIleCysThrGlnProIleProLeuLeuLysGlnLysMet 620

QY 1861 AATCATCTGTGACGATGTACAGAAACAGATCCGAGTCTTTTAGCTTAATGCTTTCTTC 1920
Db 621 AsnHisSerIleThrMetSerGlnGlnGlnIleAlaSerLeuLeuAlaAsnAlaPhePhe 640

QY 1921 TGCAATTTCCCGCAGGAATGCCAAGATGAATCGGAGTATTCTAGTATCCCGACATT 1980
Db 641 CysThrPheProArgArgAsnAlaLysMetLysSerGluTyrSerSerTrpProAspIle 660

QY 1981 AACTTCAATCGTGTGTTGAAGAGAGCTTCATCAGGAAACCCAGAAACACTGAAACACTC 2040
Db 661 AsnPheAsnArgLeuPheGluGlyArgSerArgLysProGluLysLeuLysThrLeu 680

```

```
QY 2041 TTCTGCTACTTTCGAAGAGTCACAGAGAAAAAAGCTCAGAGTTGGTGACATTTACAGA 2100
Db 681 PheCysTyrPheArgValThrGluLysPheProThrGlyLeuValThrPheThrArg 700
QY 2101 CAGAGCTTGAAGATTTCCAGATGGAAGAGGTGTGAAGAGCTCTCAGACGCTTACAC 2160
Db 701 GlnSerLeuGluAspPheProGluThrPheGluArgCysGluLysLeuLeuThrArgLeuHis 720
QY 2161 GTCACTTACAGAGGTACCATAGAGGCAACGGCGAGGATGCTACAGGTGATTTGCA 2220
Db 721 ValThrTyrGluGlyThrIleGlnGlyAsnGlyGlnGlyMetLeuGlnValAspPheAla 740
QY 2221 AATCGTTTGTGAGGPGGTGTGACTGGTCGGGACTGTGTACAGAGAAATCAGATTT 2280
Db 741 AsnArgPheValGlyGlyValThrSerAlaGlyLeuValGlnGluIleArgPhe 760
QY 2281 TTAATCAATCTGATGATTTCTTTCAGCGGTCTCACTGAGGTCTGGATCACAAGTAC 2340
Db 761 LeuIleAsnProGluLeuIleValSerArgLeuPheThrGluValLeuAspHisAsnGlu 780
QY 2341 TGTCTTATTATCACAGGTACTGAACAGTACAGTGAATACACAGGCTATGCTGAACATTAT 2400
Db 781 CysLeuIleThrGlyThrGluGlnTyrSerGluTyrThrGlyTyrAlaGluThrTyr 800
QY 2401 CGTTGGGCGCCGAGCCATGAAGATGGAGTGAAGAGCAGTATGGCGCGCGCTGCACG 2460
Db 801 ArgTrpAlaArgSerHisGluAspArgSerGluArgAspAspTrpGlnArgThrThr 820
QY 2461 GAGATCGTGCATGAGCGCACTTCACTTCAGAGCTACCTCAGATCAGTTGTCCTCTCT 2520
Db 821 GluIleValAlaIleAspAlaLeuHisPheArgArgTyrLeuAspGlnPheValProGlu 840
QY 2521 AAAGTGAGAGCTGAGCTTAAACAGCGTTACTCGCGATTCTCTCCGCTCCTGGAGTTCTCT 2580
Db 841 LysIleArgArgGluLeuAsnLysAlaTyrCysGlyPheLeuArgProGlyValSerSer 860
QY 2581 GAAATCTTCTGAGTGGCCACCGGAACTGGGGCTGTGTGCTCTTGGGGGTGACGCT 2640
Db 861 GluAsnLeuSerAlaValAlaThrGlyAsnTrpGlyCysGlyAlaPheGlyAspAla 880
QY 2641 AGATTAAGAGCTTAAATACAGATCTGGCAGCTGCTGGCGTGAACGTGAGCTGTTTAT 2700
Db 881 ArgLeuLysAlaLeuIleGlnIleLeuAlaIleValAlaValAlaGluArgAspValValTyr 900
QY 2701 TTCACCTTGGGACTCAGAGTTGATGAGAGACATTTACAGATGACACACTTCTCTTACC 2760
Db 901 PheThrPheGlyAspSerGluLeuMetArgAspIleTyrSerMetHisThrPheLeuThr 920
QY 2761 GAGAGGAGCTGGATGTGGAAAGTGTACAGTTATTTGCTTACTACTACATCAATCAAGAA 2820
Db 921 GluArgLysLeuThrValGlyGluValTyrLysLeuLeuLeuArgTyrTyrAsnGluGlu 940
QY 2821 TGCAAGAACTGTTCCACCCCTGGACACACATCAAGCTTTTATCCATTCATATACATGCT 2880
Db 941 CysArgAsnCysSerThrProGlyProAspIleLysLeuTyrProPheIleTyrHisAla 960
QY 2881 GTTCAGTCAAGTGCAGAGACCATTCAGATGCCAGATGCCAGACAGAGGCGAGC 2928
Db 961 ValGluSerCysThrGlnThrThrAsnGlnProGlyGlnArgThrGly 976
```

## RESULT 9

```
US-09-511-507-2
; Sequence 2, Application US/09511507
; Patent No. 6395543
; GENERAL INFORMATION:
; APPLICANT: JACOBSON, Myron K.
; APPLICANT: JACOBSON, Elaine L.
; APPLICANT: AME, Jean-Christophe
; APPLICANT: LIN, Winston
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE (PARG) EN
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV
; TITLE OF INVENTION: THEREWITH
; FILE REFERENCE: NIAD 201
```

```
; CURRENT APPLICATION NUMBER: US/09/511,507
; CURRENT FILING DATE: 2000-02-23
; PRIOR APPLICATION NUMBER: 09/302,812
; PRIOR FILING DATE: 1999-04-30
; NUMBER OF SEQ ID NOS: 38
; SEQ ID NO 2
; LENGTH: 977
; TYPE: PRT
; ORGANISM: Bos taurus
; FEATURE:
; US-09-511-507-2

Alignment Scores:
Pred. No.: 0 Length: 977
Score: 4332.50 Matches: 820
Percent Similarity: 89.86% Conservative: 57
Best Local Similarity: 84.02% Mismatches: 90
Query Match: 63.97% Indels: 9
DB: 4 Gaps: 5
```

US-09-302-812-5 (1-3814) x US-09-511-507-2 (1-977)

```
QY 28 ATGAGTGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 84
Db 1 MetSerAlaGlyProGlyCysGluProCysThrLysArgProArgTrpAspAlaAla 20
QY 85 ACTTCTCGCGCGCTGCTCGGACTCCCGGAGCTTCCCTGCGAGCGAGAGCGGTGTTCTC 144
Db 21 ThrSerProAlaAlaSerAspAlaArgSerPheProGlyArgGlnArgValLeu 40
QY 145 GACCCCAAGACGCTCCCGTCCAGTTCCAGGTCCTCCCTCCCTCCCTCCCTCCCTCCG 204
Db 41 AspSerLysAspAlaProValGlnPheArgValProProSerSerSerGlyCysAlaLeu 60
QY 205 GGGCGGGCGGGCCCGCACAGAGCAACGCCACCTCGTTGTTTTCACAAACAAACACTATT 264
Db 61 GlyArgAlaGlyGlnHisArgGlySerAlaThrSerLeuValPheLysGlnLysThrIle 80
QY 265 ACTACTTGGATGATCTATAAGGACCCCAAGACAGCTGAATCAGAA-----AGTAAA 315
Db 81 ThrSerTrpMetAspThrLysGlyIleLysThrValGluSerGluSerLeuHisSerLys 100
QY 316 GAAACACACATACAGAAATTTGACTCCATGATGATGTTCTGTCAGAAAGATTAATTTTAC 375
Db 101 GluAsnAsnAsnThrArgGluGluSerMetMetSerSerValGlnLysAspAspPheTyr 120
QY 376 CCACATAAGGTGAAATAATTTGAAATGTTCTCCTCAGCTAAATCTTGATAAATCAACCCACA 435
Db 121 GlnHisAsnMetGluLysLeuGluAsnValSerGlnLeuGlyPheAspLysSerProVal 140
QY 436 GAAAGAGTTTACAGTATTTGAACCAACAGCAGACTCGAGTGTGTCCAGTGGCAGAAAT 495
Db 141 GluLysGlyThrGlnTyrLeuLysGlnHisGlnThrAlaAlaMetCysLysTrpGlnAsn 160
QY 496 GAAGGGAAGCATCAGACAGACTTTTGGCAAGTGGAGCTCCCGGGGGAGCTCCGCTACCA 555
Db 161 GluGlyProHisSerGluArgLeuLeuGluSerGluProProAlaValThrLeuValPro 180
QY 556 AAGCAGCTTAGTAAATGCTAACTTGTGTGCTACCTCCCGGAGCTCCCGGAGCTCCGCTACCA 615
Db 181 GluGlnPheSerAsnAlaAsnValAspGlnSerSerProLysAspAspHisSerAspThr 200
QY 616 GATCATGAAGACAGACAGACAAATCAGACTTCTTACCTATATAAATCTTGCAATATACA 675
Db 201 AsnSerGluGluSerArgAspAsnGlnGlnPheLeuThrHisValLysLeuAlaAsnAla 220
QY 676 AAGCCACAGTACGAGATGGGAG-----GCCAGAGCAACTGTAAAGTGCAGTGGGA 726
Db 221 LysGlnThrMetGluAspGluGlnGlyArgGluAlaArgSerHisGlnLysCysGlyLys 240
QY 727 TCTCGCCGCTGTGTAAAGACTGTACAGGCTGTCAACAGCAGGAGGTGGATGTGCTACCA 786
Db 241 AlaCysHisProAlaGluAlaCysAlaGlyCysGlnGlnGluThrAspValValSer 260
```

1825 -----TGCACCCAGCCAAATACACCTCTCTGAACACAGAAAGATGAATCATCT 1869  
 Db AsnIleValSerGlyIleValThrGlyLeuArgLeuLeuAsnSerGlnGlnAlaGlyIle 138  
 1870 GTCACGATGTACAGGAACAGATGCCAGTCTTTAGCTTAATGCTTTCTTCTGACATTT 1929  
 Db ValPheLeuSerGlnGluLeuIleGlyAlaLeuLeuAlaCysSerPhePheCysLeuPhe 158  
 1930 CCCCAGCAGGAATGCCAAGATGAATCGGAGTATCTAGTTACCCAGACATTAACTCAAT 1989  
 Db ProAspAsnArgGlyAlaLys -----HisLeuProValIleAsnPheAsp 174  
 1990 CGG-----TTGTTTGAAGGAGTTTCATCAAGAAACACAGAAACAACTGAACACTCTTC 2043  
 Db HisLeuSerLeuTyrIleSerTyrSerGlnSerGlnSerLeuIleArgCysIleMet 194  
 2044 TGCTACTTTCCGAAGAGTACACAGAGAAAACCTACAGGATGGTGACATTTTACAAGACAG 2103  
 Db HisTyrPheGluArgPheCysSerCysValProIleGlyIleValSerPheGluArgLys 214  
 2104 -----AGTCTTGAAGATTTCCAGATGGGAAAGGTGTGAAGAGCTCTGACAGCTTA 2157  
 Db IleThrAlaAlaProAspAlaAspPheTyrSerLysSerAspValSerLeu ----- 231  
 2158 CACGTCACTACGAGGTTACCATAGACGAACGCCGAGGATCTACAGTGGATTTT 2217  
 Db -----TyrGlnProAsp -----AsnAlaLeuGluValAspPhe 242  
 2218 GCAATCCTTTTGTGGAGGTGTGTGCTGGCGGACTTGTACAAGAGAAATCAGA 2277  
 Db AlaAsnLysTyrLeuGlyGlySerLeuSerArgGlyCysValGlnGluGluLeuArg 262  
 2278 TTTTAAATCAATCCTGAATGATTTTTCACGGCTGTTCACTGAGGTGCTGATCACAAT 2337  
 Db PheMetIleAsnProGluLeuIleAlaGlyMetLeuPheLeuProArgMetAspAsn 282  
 2338 GAGTGTCTTATATACAGGTACTGAAACAGTACAGTGAATACACAGGCTATGCTGAAC 2397  
 Db GluAlaIleGluIleValGlyAlaGluPheSerCysTyrThrGlyTyrAlaSerSer 302  
 2398 TATCGTTGGCGCCGAAGCATCAAGATGGAGTGAAGAGGAGTGTGGCAGCGGGCTGC 2457  
 Db PheArgPheAlaGlyGluTyrIleAspLysLysAlaMetAspPheLysArgArgArg 322  
 2458 ACGGAGATCGTTGCTATGACGCTT -----CACTTCAGACGCTAC 2499  
 Db ThrArgIleValAlaIleAspAlaLeuCysThrProLysMetArgHisPheLysAspIle 342  
 2500 CTCGATCATGTTGTGCCTGAGAAAGTGAAGTGAAGAGGAGTGTGGCAGCGGGCTGC 2559  
 Db Cys -----LeuLeuArgGluIleAsnLysAlaLeuCysGlyPhe 355  
 2560 CTC ----- 2562  
 Db LeuAsnCysSerLysAlaTyrGluHisGlnAsnIlePheMetAspGluGlyAspAsnGlu 375  
 2563 -----CGTCTGGA ----- 2571  
 Db IleGlnLeuValArgAsnGlyArgAspSerGlyLeuLeuArgThrGluThrAlaSer 395  
 2572 -----GTTCTCTGAAATCTT----- 2589  
 Db HisArgThrProLeuAsnAspValGluMetAsnArgGluLysProAlaAsnAsnLeuIle 415  
 2590 -----TCTCAGAGTGGCCACG 2604  
 Db ArgAspPheTyrValGluGlyValAspAsnGluAspHisGluAspAspGlyValAlaThr 435  
 2605 GGAACCTGGGCTGTGTGCTTTGGGGTCAAGCTAGATTAAAGCCCTTAATACAGATC 2664  
 Db GlyAsnTrpGlyCysGlyValPheGlyGlyAspProGluLeuLysAlaThrIleGlnTrp 455

2665 CTGGCAGCTGTGGCGGTGAACGTGACGTGGTT-----TATTTACCTTTGGGACTCAGAG 2721  
 Db LeuAlaAlaSerGlnThrArgArgProPheIleSerTyrThrPhe-GlyValGluAl 475  
 2722 TTGATGAGAGACATTTACGACATGCACATTTTCTTCCAGAGAGGAGCTGGATTTGGA 2781  
 Db 475 a ----- 475  
 2782 AAAGTGTAACAAGTTATTCTCTAGATACATAAGAAATGAGAAATGTTTCCACCC-- 2839  
 Db -----LeuArgAsnLeuAspGlnLeuProProTh 485  
 2840 -----CTGGACACACATCAAGC-----TTTATCCATTC 2868  
 Db rLysLeuSerSerArgLeuAspSerAlaAsnSerProSerLeuThrCysLeuMetHis 505  
 2869 ATATACCATGCTGTGTAGTCAAGTGCAGAGACCACTGACATGCCAGGACAGAGCGGC 2928  
 Db rHisThrValSerLeuHis ----- 511  
 2929 ACCTGAGGAACAAGTACTAGGACCTCTCTCAAAGACATCCTATTATTGAAATGTGGG 2988  
 Db -----PheProThrTrpAl 516  
 2989 TGTGATGCTCAATGACTGAATCTGATCTAAGTGTGTATATATAATCCACATTTGTAATCA 3048  
 Db aileArgMetGluLeu--ArgAlaAspLeuArg ----- 526  
 3049 AGGATGCACTCTCTCTGATATGCAAGTGTGTCTTCTGTTTCATCTGGTGGACATGCCTTT 3108  
 Db -----SerIleLeuGlnTyrLeuProLe 534  
 3109 AGACATGCTTCTCAATTTTCTCTCTTCACTTATTTATTTATTTCTTTTCTTTTCTTC 3168  
 Db uValAlaGlnSerSerLeuValTrpProSerVal ----- 547  
 3169 AACTTGATTTCTGGGAAACTCAAGAAAGTTGCACATGCTTCTAGATCTTCTCTTC 3228  
 Db -----GluGluGluLeuGlnThrIleSerArgGlyProSerG 560  
 3229 CTGTCTGTGTGTTCTCCAGACTGCTTGTGGTGGCTAGCAGATACCATCTGGAGAA 3288  
 Db userMetValAsnSerGlyGluAlaLeu--AlaLeuHisIleThrAsnMetArgLysSe 579  
 3289 GTTACAAATCCAGAAATCTGAGTTTGTCTGCA----- 3319  
 Db rLeuSerLeuAsnAlaSerAspLeuAlaProTyrAlaLeuGlnGlyTyrGlyLeuPhePh 599  
 3320 -GATTACCTGTGAGCTTCTCA-----CTCCCAACCCCT 3351  
 Db eAspLysIleSerArgGluGluSerAlaAsnPhePheGlyGluValProAlaLe 619  
 3352 TGTTAGGCTTGTGTGTCTACATTTTCAATTTTGGAA----- 3388  
 Db uCysArgLeuLeuGlnLeuProSerMetLeuGluLysHisTyrGlnLysAlaAspHi 639  
 3388 ----- 3388  
 639 sValLeuAspGlyValLysSerGlyLeuArgLeuLeuGlyProGlnGluAlaGlyIleVa 659  
 3389 -----GTTCAAGTCTTTCTTATGTTACTTAAATGCTAGTATCTTTAGGCT----- 3433  
 Db LeuLeuSerGlnGluLeuIleAlaLeuLeuAlaCysSerPhePheCysLeuPhePr 679  
 3434 -----AAACTATTTTCTATTAAAGGACAGACTAAATTTCCAGTTTCTCTTTTGGAA 3483  
 Db oGluValAspArgSerLeuLysAsnLeuGlnGlyIleAsnPheSerGlyLeuPhe----- 697  
 3484 CATCATCCCTATTAAGTAACGGTTTTTTCCTCTTTTCCCGACGCTATTTTAGAAGC 3543  
 Db -----SerPheProTyrMetArgHisCysTh 706  
 3544 TGGCCAGAGGAAGA-----AAATGTAGATAAAGGATTTTCTCGGATCTATAAGAA 3600



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: May 26, 2004, 16:16:17 ; Search time 55.3609 Seconds  
(without alignments)  
13253.917 Million cell updates/sec

Title: US-09-302-812-5  
Perfect score: 6773  
Sequence: 1 999999actgtgtctgcggg.....atcatttttcagaaaaaa 3814

Scoring table: BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 566732

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame\_n2p.model -DEV=xlp  
-Q/cgn2.1/USPTO\_spool\_p/6333148/runat 26052004 150054 5689/app query.fasta\_1.12437  
-DB=PIR\_78 -QFMT=fastan -SUFFIX=n2p.rpr -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOCALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=6333148 @CGN 1 1 298 @runat 26052004 150054 5689 -NCPU=6 -ICPU=3  
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : PIR\_78.\*  
1: pir1.\*  
2: pir2.\*  
3: pir3.\*  
4: pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	539.5	8.0	997	B84726	probable poly(ADP-ribose) glycohydrolase [imported] - Arabidopsis thaliana
2	443.5	6.5	726	T21138	hypothetical prote
3	328	4.8	364	A84726	probable poly(ADP-ribose) glycohydrolase [imported] - Arabidopsis thaliana
4	176	2.6	6642	T29757	protein UNC-89 - C
5	171	2.5	997	T43523	cut17 protein - fi
6	169	2.5	1165	T16420	hypothetical prote
7	166.5	2.5	699	I38073	nucleolar phosphop
8	166	2.5	1210	I39410	AF-4 protein, spli
9	165	2.4	1192	T71623	probable secreted
10	163	2.4	532	T96029	hypothetical prote
11	163	2.4	2722	T20532	hypothetical prote
12	163	2.4	2738	T88320	protein F07A11.6 (
13	162	2.4	1213	A58198	serine/proline-ric
14	161.5	2.4	773	F90537	lipoprotein impor

#### ALIGNMENTS

##### RESULT 1

B84726

probable poly(ADP-ribose) glycohydrolase [imported] - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 02-Feb-2001  
C:Accession: B84726  
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.; I  
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J  
Nature 402, 761-768, 1999  
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.  
A:Reference number: B84420; MUID:20083487; PMID:10617197  
A:Accession: B84726  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-997 <STO>  
A:Cross-references: GB:AE002093; NID:G4887749; PIDN:AAD32285.1; GSPDB:GN00139  
C:Genetics:  
A:Gene: At2g31870  
A:Map position: 2

##### Alignment Scores:

Pred. No.:	9,41e-30	Length:	997
Score:	539.50	Matches:	205
Percent Similarity:	36.62%	Conservative:	88
Best Local Similarity:	25.62%	Mismatches:	213
Query Match:	7.97%	Indels:	294
DB:	2	Gaps:	27

US-09-302-812-5 (1-3814) x B84726 (1-997)

QY	1732	TGGGATAGGTACTTGAAGACGACGCCAACATTATATCATCTTACCTGAC	1791
Db	79	PheAspGluLeuLeuAspGluLysGluSerLysArgTrpPheAspGluLeuLeuProAla	98
QY	1792	ATGCTGAAATTCACCTCTCTCTCCCAATATT	1824
Db	99	LeuAlaSerLeuLeuLeuGluInPheProSerLeuLeuGluValHisPheGlnAsnAlaAsp	118

natural killer cel  
nucleolar phosphop  
hypothetical prote  
hypothetical prote  
ENL (translocation  
hypothetical prote  
coagulation factor  
hypothetical prote  
enamelin matrix pr  
PAR18-1 protein -  
AAS surface protei  
DNA helicase YGI15  
supervillin P205 -  
hypothetical prote  
proliferation pote  
hypothetical prote  
microtubule-associ  
SRP40 protein - ye  
ras-responsive ele  
dentin matrix acid  
protein B0464.5a l  
TCOP1 protein - mo  
hypothetical prote  
Duffy receptor - p  
dextranase - Strep  
DNA-directed RNA p  
membrane nuclease,  
beta-spectrin - mo  
spectrin beta chai  
hypothetical prote  
natural killer cel

15	160.5	2.4	1403	1	A47328
16	159.5	2.4	990	2	T51618
17	159.5	2.4	1001	2	T16419
18	159.5	2.4	1115	2	T19437
19	157.5	2.3	559	2	B44265
20	157.5	2.3	1359	2	T34036
21	157.5	2.3	2211	1	KFB05
22	157.5	2.3	2526	2	T20531
23	156	2.3	1274	2	T37193
24	155.5	2.3	917	2	JC7799
25	155	2.3	1463	2	T30290
26	155	2.3	1489	2	S60416
27	155	2.3	1792	2	T08878
28	155	2.3	2361	2	T25752
29	154.5	2.3	1560	2	T42727
30	153	2.3	1611	2	T38236
31	153	2.3	5327	2	T13564
32	152.5	2.3	406	2	S38170
33	152.5	2.3	1615	2	JC6510
34	152	2.2	489	2	A45988
35	151	2.2	1093	2	F88556
36	151	2.2	1320	2	JC5630
37	151	2.2	1791	2	T02345
38	150	2.2	1070	2	T30848
39	150	2.2	1337	2	T30291
40	150	2.2	2910	2	T28156
41	149.5	2.2	1125	2	E90598
42	149.5	2.2	2128	2	I52577
43	149.5	2.2	2137	1	SUH0B
44	149	2.2	1401	2	T48079
45	149	2.2	1507	2	B47328



Job time : 248.584 secs

```
Db      785 sAlaGluAspLysGlnProAlaGlyLysHisSerSerGluLysArgSerSerAs 805
QY      1837 -----ATACACTCTCGAACAAGAGATG-----AATCA 1865
Db      805 pSerSerLysLeuAlaLysLysArgLysGlyGluAlaGluArgAspCysAspAsnLy 825
QY      1866 TTCCTCAGATG-----TCACAGGAACAGATCGCCAGTCTTTAGCTAA 1910
Db      825 sLysIleArgLeuGluLysGluIleLysSerGlnSerSerSerSerHisLy 845
QY      1911 TGTCTTCTCTCCACATTTCCCGACGGAATGCCAAGATGAAATCGAGATATTCTAGTTA 1970
Db      845 sGluSerSerLysThrLysProSerArgProSerSerGlnSerSerLysGluMetIe 865
QY      1971 CCCAGACATTAACCTCAATCGGTTCTTTGAAGACGTTTCATCAAGAAACAGAAAAA-- 2028
Db      865 uProProProValSer-----SerSerSerGlnLysProAlaLysPr 880
QY      2029 ----CTGAAACACTCTTCTGTACTTTTCGAGAGTCACAG-----AAAAACC 2075
Db      880 oAlaLeuLysArg-----SerArgArgGluAlaAspThrCysGlyGlnAspPr 896
QY      2076 TACAGGATTGCTGACATTACAAGACAGAGTCTTCAAGAT-----TTTCAGAAATGGGA 2129
Db      896 oProLysSerAlaSerSerThrLysSerAsnHisLysAspSerSerIleProLysGlnAr 916
QY      2130 AAGGTGTAAAGGCTCTGACACGCTTACACGCTTACGCTCTTACGAGGTTACCAT-- 2181
Db      916 gArgValGluGlyLysGlySerArgSerSerGluHisLysGlySerSerGlyAspTh 936
QY      2182 -----GAGGGCAACGGCGA-----GGCATGCTTACA 2207
Db      936 rAlaAsnProPheProValProSerLeuProAsnGlyAsnSerLysProGlyLysProGl 956
QY      2208 GGTGGATTTTCAAAATCGTTTGTGTGGAGTGGTGTGACTGCTGGTGGGACTT---GTACA 2264
Db      956 nValLysPheAspLysGln-----GlnAlaAspLeuHisMetAr 969
QY      2265 AGAAGAAATCAGATTTTAAATCAATCTGAATGATTTTCACGGCTG----- 2313
Db      969 gGluAlaLysLysMetLysGlnLysAlaGluLeuMetThrAspArgValGlyLysAlaph 989
QY      2314 -----TTCACAGCTGCTGATCAATGAGTGTCTTATATATACAGGTTACTGAC 2366
Db      989 eLysTyLeuGluAlaValLeuSerPheIleGluCysGlyIleAlaThrGluSerGlus 1009
QY      2367 GTACAGTCAA-----TACACAGCTATGCTGAAACT-----TA 2399
Db      1009 rGlnSerSerLysSerAlaTySerValTySerGluThrValAspLeuIleLysPheIl 1029
QY      2400 TCGTTGGCCCGAAGCCATGAAGATGGAGTGAAAGACGATTGGCAG----- 2448
Db      1029 eMetSerLeuLysSerPheSerAspAlaThrAlaProThrGlnGluLysIlePheAlava 1049
QY      2449 -----CGGCGCTGCACGAGATCGTTGCCATTCAGCAGCTTCACCTTCAGACGCTACCT 2501
Db      1049 lLeuCysMetArgCysGlnSerIle-----LeuAsnMetAlaMetPheArgCysLysLy 1067
QY      2502 CGATCAGTTTGTGCTGAGAAAGTGAGACGTGAGCTTAACAAGGCTTACTCGCGGATTCCCT 2561
Db      1067 sAsp-----IleAlaIleLysTySerArgThrLeuAsnLysHisPhe----- 1081
QY      2562 CCGTCTGGAGTTCCTTCTGAAATCTTTCTGCAAGTGGCCACGGGAAACTGGGGCTGTGG 2621
Db      1082 -----GluSerSerSerLysValAlaGlnAlaProSerProCysIl 1095
QY      2622 TGCCTTTGGGGT 2634
Db      1095 eAlaSerThrGly 1099
```

Search completed: May 26, 2004, 18:37:57

15 TC CGG-----TCC CAGCATGAGTCGGCGCCCGCGCTG-----GGA 50  
151 CysGlyProProAspSerGlnHisLeuThrGlnAspArgLeuGlyGlnGluGlyPheGly 170  
51 GCCCTGCACGAAGCGCGCTGGGGCGCCCTGGAACCTCTGCGCCGACCTGCCTCGACTC 110  
171 SerSerHisHisLeuGlyGlyAspArgArg-AlaAspGlyAspHisCysAlaSerValTh 190  
111 CGGAGCTTCCCTGGCAGGCAGAGCGCTGTTCTCCGACCCCAAGGACGCTCCCGTCCAGTT 170  
190 rAspSerAlaProGluArgGluLeuSerProLeuLeuSerLeuProSerPro----- 207  
171 CAGGTCCTCTCGTCTCCGACGCTGCTGCTGCGGGCGGGCGGACCGCAGAGGCAA 230  
208 ----ValProProLeuSerProIleHisSerAsnGlnGlnThrLeuProArgThrGlnG 226  
231 GCGCACCTCGTTGTTTTCACAAACAAGACTATTACTCTTGATGGATACATAAAGGACC 290  
226 ySerSerHisValHisGlySerSerAsnAsnSerHisGlyTyrCysProAlaLeuSerP 246  
291 CAAG-----ACAGCTGAATCAGAAAGTAAGAAACAAACAATACAAGAATTCAGCTCCAT 344  
246 OlyAspLeuAlaValLeuValHisAspLysGlu-----ThrProGlnAspSerLe 263  
345 GATGAGTTCTGTGCAGAAAGATAAC-----TTTTACCCCAATAAGGT 386  
263 uValAlaProAlaGlnProProSerGlnThrPheProProProSerLeuProSerLys 283  
387 GGAATAATTTGGAATAATGTTCTCAGCTAAATCTT-----GATAATC 428  
283 rValAlaMetGlnGlnLysProThrAlaTy rValArgProMetAspGlyGlnAspGlnAl 303  
429 ACCCAGAGAAAG-----AGTTTCAGATATTTGAACCAACACAGACAG 470  
303 aProSerGluSerProGluLeuLysProLeuProGluAspTyr-----ArgGlnGlnTh 321  
471 T-----CCGAGTGTGTCCAGTGTCGACGAGTGA 497  
321 rPheGluLysThrAspLeuLysValProAlaLysAlaLysLeuThrLysLeuLysMetPr 341  
498 AGGGAAGCATCAGAAACAGCTTTTGGCAAGCAGCCTCCCGGGGACCTCCGCTACCAA 557  
341 oSerGlnSerValGluGlnThrTyrSerAsnGluValHisCysValGluGluLeuLeu 361  
558 GCAGCTTAGTAATGCTAACATGCTGCTCAGTCCACCCACACTGATGACACAGTGCACAGA 617  
361 sGluMetThrHisSer-----TrpProProLeuThrAlaIleHisThrProSerTh 379  
618 TCATGAAGACAGACAGACATCAGAGTTCTTACACTATATAAACTTTGAAATACAAA 677  
379 rAlaGluProSerLys-----PheProPheProThrLysAspSerGlnHisVa 395  
678 GCCAACAGTAGGAGATGGCAGGCGCAGGAAGCACTGTAAGTGCAGTGGATCTGCCAGTC 737  
395 lSerSerValThrGlnAsnGlnLysGlnTyrAsp-----ThrSe 408  
738 TGTGAAAGACTGTGACAGCTGTCAACAGGAGGAGGTGGATGTGTCTACACAGAGT---CC 794  
408 rSerLysThrHisSerAsnSerGlnGlnGlyThrSerSerMetLeuGluAspAspLeuGl 428  
795 TTTGTCAGATTTGGTCCCGAGGACATTGGAACTGGACCAAAAAATGACAAACAATTGAC 854  
428 nLeuSerAsp-----SerGluAsp-----SerAspSerGluGlnTh 440  
855 TGGCAAGAAGACAGCCTAGGTGATTGCGCTCCATTTTCAGAAAGAAAGTGGAGCTGTGTC 914  
440 rProGluLysProProSerSerSerAlaProProSerAlaProGlnSerLeuProGlu- 459  
915 ACCAATGGATGTAGACAACTCGAGAAACAGTTGTCAAGATTTCAGAGCAGATGAAGAAAC 974  
460 -ProValAlaSerAlaHisSerSerSerAlaGluSerGluSerThrSerAspSerAspSe 479

[illegible]

QY 1519 TGG---GATAACAAACATGTGAATGCTTGTTCGGAACAAACATGTGTACCTGTGGAA 1575  
Db 729 TrpIleAspAsnLys 733  
QY 1576 GATGAGATGGTGCAGGAACTGCAGGAGTGTAGTGGAGTCATTTCAGACTGCACCTTCTC 1635  
Db 733 733  
QY 1636 AACAAATTTCACACGACCCAGCACTTTGAAGATGCGAATCTGAATPACAAATGTGGCATAT 1695  
Db 734 734  
QY 1696 TCTAAGAAATGGCACTTTACAGCTTTGGTTGATTTCTGGGATGAAGTACTTGAAGAGCA 1755  
Db 747 747  
QY 1756 GAGGCCCAACATTTATATATCATGTCCTTATTCCTGACATGCTGAAATTCGACTCTGTCTG 1815  
Db 749 GluValLysAsnValThrGluGly 756  
QY 1816 CCAATATTTGACCCAGCCCAATACACCTCTGGAACAGAGATGAATCTGTCTGACG 1875  
Db 757 ProSerValThrArgPheGluLeuSerValGluLysGlyValLysValSerArgIleThr 776  
QY 1876 ATGTCACAGAACAGATCCCGAGTCTTTTAGCTAATGCTTTCTGTCACATTTCCCGA 1935  
Db 777 AlaLeuGlnAspAspLysMetAlaLeuAla 787  
QY 1936 CGGAATCCCAAGATGAATCGGAGTATTCTAGTTACCCAGCAATTAATCTCAATCGGTG 1995  
Db 788 788  
QY 1996 TTTGAAGACGTTCTATCAAGGAACACAGAAACACTGTAACACTCTTCTGCTACTTTGCA 2055  
Db 797 IleProGlyThrSerLeuValGlyIleGlu 806  
QY 2056 AGAGTCACAGAAACACCTACAGGATTTGGTGCATTTACAGACAGAGTCTTGAAGAT 2115  
Db 807 807  
QY 2116 TTTCCAGATGGGAAGGTGTGAAGGCTCTGACACGCTTACAGCTCACTTACGAGGT 2175  
Db 823 SerProLysPheLysAsnThrGluSerLysLeuThr 838  
QY 2176 ACCATAGAACCAAGCCGAGGATGTCTACAGGTGGATTTGCAATCGTTTGTGGA 2235  
Db 839 TyrArgIleAsnAsnGluProLeuLeuMetAspIleAlaLysThrProHisAlaLeuIle 858  
QY 2236 GGTGTGTGACTGTGGCGGACTTGTGA 2262  
Db 859 AlaGlyAlaThrGlySerGlyLysSerValCysIleAsnSerIleLeuMetSerLeuLeu 878  
QY 2263 2263  
Db 879 TyrLysAsnHisProGluGluLeuArgLeuLeuIleAspProLysMetValGluLeu 898  
QY 2302 2302  
Db 899 AlaProTyrAsnAspLeuProHisLeuValSerProValIleThrAspValLysAlaAla 918  
QY 2335 AATGAGTGTCTTATATTCACAGGTACTGACAGGTACAGTGAATACACAGGCTATGCTGAA 2394  
Db 919 ThrGlnSerLeuLysTrpAlaValGluGluMetGluLysArgTyrLysLeuPheAlaGln 938  
QY 2395 ACTTATCGTTGGCCGAGGACCATGAAGATGGAGTGAAGAGGAGTGGCAGCGGC 2454  
Db 939 TyrHis 956  
QY 2455 TGCACGAGATCGTT---GCCATTGACGCACTT---CACITTCACAGCTTACCTGCATCAG 2508  
Db 957 MetProLysIleValIleValIleAspGluLeuAlaAspLeuMetMetAlaProGln 976

QY 2509 TTTGTGCTCGACAAAGTAGAGCTGAGCTTACAGGCT--- 2547  
Db 977 AspValGluGlnSerIleAlaAlaArgIleAlaGlnLysAlaArgAlaCysGlyIleHisMet 996  
QY 2548 2548  
Db 997 LeuValAlaThrGlnArgProSerValAsnValIleThrGlyLeuIleLysAlaAsnIle 1016  
QY 2575 CTTCTCGAAATCTTTCTGTCAGTGGCCACCGGAACTGGGGCTGTGGTGCCTTTGGGGGT 2634  
Db 1017 ProThrArgIleAlaPheMetValSerSer 2688  
QY 2635 GACGCTAGATTAAGAGCTTAAATACAGATCTCGGAGCTGTGGCGCTGAAGCT--- 2688  
Db 1029 AspSerArg---ThrIleLeuAspSerGlyGlyAlaGluArgLeu 1043  
QY 2689 2689  
Db 1044 GlyTyrGlyAspMetLeuTyrLeuGlySerGly---MetAsnLysProIleArg 1060  
QY 2740 AGCATGCACATTCCTTACCGAGAGGAAGCTGGAT 2775  
Db 1061 ValGlnGlyThrPheValSerAspAspGluIleAsp 1072

RESULT 15  
US-10-205-823-264  
; Sequence 264, Application US/10205823  
; Publication No. US20030108963A1  
; GENERAL INFORMATION:  
; APPLICANT: Schlegel, Robert  
; APPLICANT: Monahan, John E.  
; APPLICANT: Endege, Wilson O.  
; APPLICANT: Gannavarapu, Manjula  
; APPLICANT: Gorbacheva, Bella  
; APPLICANT: Hoersch, Sebastian  
; APPLICANT: Kamatkar, Shubhangi  
; APPLICANT: Womsey, Angela M.  
; APPLICANT: Glatt, Karen  
; APPLICANT: Zhao, Xumei  
; APPLICANT: Anderson, Dustin  
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND  
; METHODS FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND  
; THERAPY OF PROSTATE CANCER  
; FILE REFERENCE: MRI-044  
; CURRENT APPLICATION NUMBER: US/10/205,823  
; PRIOR FILING DATE: 2002-07-25  
; PRIOR APPLICATION NUMBER: 60/307,982  
; PRIOR FILING DATE: 2001-07-25  
; PRIOR APPLICATION NUMBER: 60/314,356  
; PRIOR FILING DATE: 2001-08-22  
; PRIOR APPLICATION NUMBER: 60/325,020  
; PRIOR FILING DATE: 2001-09-25  
; PRIOR APPLICATION NUMBER: 60/341,746  
; PRIOR FILING DATE: 2001-12-12  
; PRIOR APPLICATION NUMBER: 60/362,158  
; NUMBER OF SEQ ID NOS: 455  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 264  
; LENGTH: 1210  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-205-823-264

Alignment Scores:  
Pred. No.: 0.000451 Length: 1210  
Score: 166.00 Matches: 216  
Percent Similarity: 33.43% Conservative: 140  
Best Local Similarity: 20.28% Mismatches: 402  
Query Match: 2.45% Indels: 307  
DB: 14 Gaps: 48

US-09-302-812-5 (1-3814) x US-10-205-823-264 (1-1210)

```
/ APPLICANT: Ohlsen, Kari
/ APPLICANT: Zyskind, Judith
/ APPLICANT: Wall, Daniel
/ APPLICANT: Trawick, John
/ APPLICANT: Carr, Grant
/ APPLICANT: Yamamoto, Robert
/ APPLICANT: Forsyth, R.
/ APPLICANT: Xu, H.
/ TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
/ FILE REFERENCE: ELITRA.034A
/ CURRENT APPLICATION NUMBER: US/10/282,122A
/ PRIOR FILING DATE: 2003-02-20
/ PRIOR APPLICATION NUMBER: 60/191,078
/ PRIOR FILING DATE: 2000-03-21
/ PRIOR APPLICATION NUMBER: 60/206,848
/ PRIOR FILING DATE: 2000-05-23
/ PRIOR APPLICATION NUMBER: 60/207,727
/ PRIOR FILING DATE: 2000-05-26
/ PRIOR APPLICATION NUMBER: 60/230,335
/ PRIOR FILING DATE: 2000-09-06
/ PRIOR APPLICATION NUMBER: 60/230,347
/ PRIOR FILING DATE: 2000-09-09
/ PRIOR APPLICATION NUMBER: 60/242,578
/ PRIOR FILING DATE: 2000-10-23
/ PRIOR APPLICATION NUMBER: 60/253,625
/ PRIOR FILING DATE: 2000-11-27
/ PRIOR APPLICATION NUMBER: 60/257,931
/ PRIOR FILING DATE: 2000-12-22
/ PRIOR APPLICATION NUMBER: 60/267,636
/ PRIOR FILING DATE: 2001-02-09
/ PRIOR APPLICATION NUMBER: 60/269,308
/ PRIOR FILING DATE: 2001-02-16
/ Remaining Prior Application data removed - See File Wrapper or PALM.
/ NUMBER OF SEQ ID NOS: 78614
/ SOFTWARE: PatentIn version 3.1
/ SEQ ID NO 71179
/ LENGTH: 1169
/ TYPE: PRT
/ ORGANISM: Staphylococcus epidermidis
US-10-282-122A-71179

Alignment Scores:
Pred. No.: 0.00369 Length: 1169
Score: 167.00 Matches: 182
Percent Similarity: 32.56% Conservative: 128
Best Local Similarity: 19.12% Mismatches: 290
Query Match: 2.47% Indels: 352
DB: 12 Gaps: 44

US-09-302-812-5 (1-3814) x US-10-282-122A-71179 (1-1169)
QY 295 ACAGCTGAATCAGAAAGTAAGAAACAAACAATACAGAAATGACTCCATGATGAGTTCT 354
D 348 ThrGlyGluSerSerLeuAspLeuGluAsnGluSerAsnGlnAspSerSerAsnSer 367
QY 355 GTGCAGAAA-----GATACTTTTACCACATAGTGGAAAAA 393
D 368 LeuGluLysGlnSerAsnSerSerAsnLeuAspAsnLysGluAlaGlnAsnAsnThrPro 387
QY 394 TTGAAAATGTTCTCCAGCTAAATCTTGATAAATCACCACAGAAAAGAGTTTCACAGTAT 453
D 388 LeuPheAsnTyrGluGluLeuAspLeuAsp-----ThrThrSerAspValTyrLys 404
QY 454 TTGAACCAACAGCAGACTCGGAGTGTGTGCAAGTGGCAGAAATGAAGGGAAGCATGCAGAA 513
D 405 ValAsnGluGluGluThr-----GluSerLysAsnAspGlu 416
QY 514 CAGCTTTTGGCAAGTGGAGCTCCCGGGGACTCCGCTACCAACAGCAGCTTAGTAATGCT 573
D 417 AspLeuValSerSer-----CAT 621
QY 574 AACATTGGTCAGTCAACCCACACACTGATGACACACAGTACACAGAT-----CAT 621
D 716 ---LeuLeu-----GluAlaProGluPro-----HisGluLysAspGlnAsp 728
```

```
Db 422 -----AsnHisTyrHisSerAsnAspAspAlaGluValGluAspAlaGluTyrHis 438
QY 622 GAAGAAGACAGACAGCAATCAGACAGATTCTTACACCTATATAAACTGCAATAACAAG--- 678
D 439 GluLeuAspAsnArgGlnGln-----AsnGlnSerAsnSerGlnAsp 453
QY 679 CCAACAGTAGAGATGGCGAGCCAGAACAACTGTAAGTCAGTCAGTATCCCGAGTCT 738
D 454 AspIleSerSerLysSerSerThrSerAsnMetTyrAspAsnAlaIleSerAlaSer 473
QY 739 GTG-----AAAGACTGTACAGGCTGTCAACAGGAG 768
D 474 ValAspAsnAsnThrGluArgAlaLysSerAsnGluAspLysAsnAspThrGluIleThr 493
QY 769 GAGGTGATGTCTACACAGAGATCTTTGTTCAGATGTTGGTCCCGAG----- 816
D 494 HisLeuAspGlyThrThrSerAlaLysValSerAspGluLysIleGluSerAsnThrAsn 513
QY 817 -----GACATGGAACTGGACCAAAAAATGACAACAATG----- 852
D 514 AsnHisLeuGluGlnAspLysAsnValLysLeuLysAsnValAsnSerLeuLysSerSer 533
QY 853 -----ACTGGACAA-----GAAAGCAGCTAGTGTGATTCGCCTCATTT----- 891
D 534 AsnSerAspThrGlyGlnThrArgLysGlnA-gPheGlyLysArgProPheAsnVal 553
QY 892 -----CAGAAAGAAAGTGTAG 906
D 554 LeuMetThrProSerAspLysLysArgMetMetAspGlnAsnHisLysLysValSerVal 573
QY 907 CTTGATCAACCAATGATGTAGCAACTGCGAAGAACAGTTGTCAAGATTCAGAAAGCAGAT 966
D 574 ProGluLeuLysProGluLysGlnAlaAsnHisArgLysAspSerGluSerAsn 593
QY 967 -----GAAGAAACAGTCCAGTCTTTGAT 990
D 594 LysSerGluGluPheLysGlnIleAsnThrAsnA-gGluThrAspSerAsnSerTyrGlu 613
QY 991 GAG-----CAAGATGATCGTTCTCCCAACAGCAATAATAACTTTCAAGTTGCCAA 1041
D 614 SerAsnGlyIleGluHisAspMetAsnSerSerSerAspGluHisValTyrGluThrPro 633
QY 1042 GCAAGAGAGCTGATGGCGATCTTAGGAAACGGTATTGACTTAAGGGAAGTAGTGAAGTTAGA 1101
D 634 SerLysGlnGlnAspGluGlnIleGlnLys-----LeuGln 645
QY 1102 TTGCATTTCATTTGAAGGAAATAATATCTGGGACCAGTCGACTTAATCCCAAGCCA 1161
D 646 AspAspPheHisPheGluAsnAlaAsnHisAlaLysIleAsn-----AsnSerAsnGlu 663
QY 1162 TCTGAAACTCTTCTAGCCTTAATGTAGAGTGTAGAGTTCCAAGCAGCATCGAAAAAGG 1221
D 664 ThrGlyAsnGlnSerAsnIleSer-----HisSerLysArg 675
QY 1222 GATTTCTAAATTCAGATCATTTTCATGAGAATTTCCAGTCAGAGCAGCAAGAAAGAA 1281
D 676 SerGlnTyrSerThrAsnGlu-----SerLysAsnIleAspThrGlnThrSer 691
QY 1282 CAATGTGAGTCAGACATCAAGAACACAAAGGAGATTCCAAAATACATCCACCTAAC 1341
D 692 AsnSerSerThrSerAsnGlnAsnPheGlnArg----- 702
QY 1342 CTCCTCCAGAGAAGAGTGGCTGGAACTCCTATTGAGGAATAAGAAAAATGCTCGG 1401
D 703 -----IleArgLysGlyPro----- 707
QY 1402 TGTGGATTCATTTGCTTCTTCTTAAGCATTCTGCAAGTCACACAGTGCATGTTTGGGTA 1461
D 708 ---AsnIleLysLeuProSerTyrGln----- 715
QY 1462 GACCTTCTGAGCAGGAGGAGGTTCCGAAACCTTTTCCACACATTACAAAGATTG--- 1518
D 716 ---LeuLeu-----GluAlaProGluPro-----HisGluLysAspGlnAsp 728
```

```

Db      813 ValThrSerThrLysAsnHisLysAspGlyAspAlaAsnHis----- 827
QY      1783 TTACTGACATGGTCAAAATTCGACTCTGTCTG---CCAAATATTTCACCCAGCCCAATA 1839
Db      828 -----LysAlaProIleArgGlyProAsnLysLeuProSerLeu 842
QY      1840 CCATCTCTGAACAGAGATCAATCATCTCTCAGATGTCACAG---GAACAGATGCC 1896
Db      843 AsnLeuLeuGluAspHisGluGluHisGluLysLeuGluLysLeuGluLysLys 862
QY      1897 AGTCTTTAGTAAATGCTCTCTCTGACATTCCTCCGACGGAATGCCAAGATGAATCG 1956
Db      863 GluGluLeuAsnAspAlaPheTyr----- 870
QY      1957 GAGTATTCTAGTACCCAGACATTAATCAATCGTTGTTGAGAGAGCTTCATCAAGG 2016
Db      871 ---TyrPheAsnValPro---AlaGluValGlnAsnValThrGluGlyProSerValThr 888
QY      2017 AAACCCAGAAAACCTGMAAACACTCTCTGCTACTTTCGAGAGTCACAGAGAAAACCT 2076
Db      889 Arg-----PheGluLeuAlaValGluLys----- 896
QY      2077 ACAGATTGTCACATTACAGACAGAGTCTGAAGATTTCAGAAATGGGAA----- 2130
Db      897 ---GlyValLysValSerArgIleThrAlaLeuGlnAspIleLysMetAlaLeuAla 915
QY      2131 -----AGGTGTGAAACCT----- 2145
Db      916 AlaLysAspIleArgIleGluAlaProIleProGlyThrSerLeuValGlyLeuVal 935
QY      2146 ---CTGACACGCTTACACGCTCACTTACGAGGTACCATAGAGGC----- 2187
Db      936 ProAsnLeuAsnProThrLysValAsnLeuLysSerIleLeuGluSerProLysPheLys 955
QY      2188 AACGCCGAGCGATGCTACAGTGGATTTTCGAATCGTTT----- 2229
Db      956 AsnAlaGluSerLysLeuThrValAlaMetGlyAsnArgIleAsnAsnGluProLeuLeu 975
QY      2230 -----GTTGGAGGTGCTGACTGTGCTGCGGACTGTGTA 2262
Db      976 MetAspIleAlaLysThrProHisAlaLeuIleAlaGlyAlaThrGlySerGlyLysSer 995
QY      2263 -----CAAGAAGAATCAGA 2277
Db      996 ValCysIleAsnSerIleLeuMetSerLeuLeuTyrLysAsnHisProGluGluLeuArg 1015
QY      2278 ---TTTTTAATCAATCTCGAATTGATT----- 2301
Db      1016 LeuLeuLeuIleAspProLysMetValGluLeuAlaProTyrAsnAspLeuProHisLeu 1035
QY      2302 GTTTCACGGCTGTTCACTGAGGTGCTGGATCACAATGAGTGTCTTATTATCAGAGTACT 2361
Db      1036 ValSerProValIleThrAspValLysAlaAlaThrGlnSerLeuLysTrpAlaValAsp 1055
QY      2362 GAACAGTACAGTGAATACACAGGCTATGCTGAACATTCTGTTGGGCCCAAGCCATGAA 2421
Db      1056 GluMetGluLysArgTyrLysLeuPheAlaGlnPheHis-----ValArgAsnIleThr 1073
QY      2422 GATGGAGTGAAGAAGCAGATGGCAGCGCTGCACGGAGATCGTT---GCCATTGAC 2478
Db      1074 AlaPheAsnLysLysAlaProTyrGluGlnArgMetProLysIleValIleValIleAsp 1093
QY      2479 GCACCT---CACCTCAGACGCTACCTCGATCAGTTGCTGCTGAGAAAGTGACGCTGAG 2535
Db      1094 GluLeuAlaAspLeuMetMetAlaProGlnGluValGluGlnSerIleAlaArgIle 1113
QY      2536 CTTTAAACAGGCT----- 2547
Db      1114 AlaGlnLysAlaArgAlaCysGlyIleHisMetLeuValAlaThrGlnArgProSerVal 1133
QY      2548 -----TACTGGCGATTCCTCGCTCGGAGTCTCTCTGAAAATCTTCTCGAGTGGCC 2601

```

```

Db      1134 AsnValIleThrGlyLeuIleLysSerAsnIleProThrArgIleAlaPheMetValSer 1153
QY      2602 ACGGAAACTGGGCTGTGTGCTTTGGGGGTGACGCTAGATTAAAGCCTTAATACAG 2661
Db      1154 Ser-----SerValAspSerArg-----Thr 1160
QY      2662 ATCCGTGCAGCTGCTGGCTGACGT-----GACGTGGTTTATTTCACC 2706
Db      1161 IleLeuAspSerGlyGlyAlaGluArgLeuLeuGlyTyrGlyAspMetLeuTyrLeuGly 1180
QY      2707 TTTGGGCACTCAGAGTTGATGAGACATTTACAGCATGACACTTTCCTTACCGAGAGG 2766
Db      1181 SerGly-----MetAsnLysProIleArgValGlnGlyThrPheValSerAspAsp 1197
QY      2767 AAGCTGGAT 2775
Db      1198 GluIleAsp 1200

RESULT 13
US-09-973-451-11
; Sequence 11, Application US/09973451
; Patent No. US20020132328A1
; GENERAL INFORMATION:
; APPLICANT: JACOBSON, Myron K.
; APPLICANT: JACOBSON, Elaine L.
; APPLICANT: AM. Jean-Christophe
; APPLICANT: LIN, Winston
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE
; TITLE OF INVENTION: (PARG) ENZYMES,
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIVE
; TITLE OF INVENTION: THEREWITH
; FILE REFERENCE: NIAID 201
; CURRENT APPLICATION NUMBER: US/09/973,451
; CURRENT FILING DATE: 2001-10-09
; PRIOR APPLICATION NUMBER: US/09/302,812
; PRIOR FILING DATE: 1999-04-30
; PRIOR APPLICATION NUMBER: 60/083,768
; PRIOR FILING DATE: 1998-05-01
; NUMBER OF SEQ ID NOS: 38
; SEQ ID NO 11
; LENGTH: 31
; TYPE: PRT
; ORGANISM: Bos taurus
; FEATURE:
US-09-973-451-11

Alignment Scores:
Pred. No.: 7,6e-05 Length: 31
Score: 167.00 Matches: 31
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 2.47% Indels: 0
DB: 9 Gaps: 0

US-09-302-812-5 (1-3814) x US-09-973-451-11 (1-31)

QY      2311 CTGTTCACTGAGTGTGGATCAGATGATGCTTATTATCAGAGTACTGACAGTAC 2370
Db      1 LeuPheThrGluValLeuAspHisAsnGluCysLeuIleIleThrGlyThrGluGlnTyr 20
QY      2371 AGTGAATACACAGGCTATGCTGAACCTTATCGT 2403
Db      21 SerGluTyrThrGlyTyrAlaGluThrTyrArg 31

RESULT 14
US-10-282-122A-71179
; Sequence 71179, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert

```

QY	868	AGCCTAGGTGATTCCGCTCCCATTTTGAGAAAGAAAGTGCAGCTGCAGTACCAATGGATGTA	927
Db	484	GlulysPheAspAspTyrLeuGluAspThrAsnSerHisGluGluMetLeuHisAspAsp	503
QY	928	GACAACTCGAGAAACAGCTTGTCAGATTTCAGACGACATCAAGAAACAACTGCAGTCTTT	987
Db	504	AspLeuHisGluGlnValMetAspAspAspGluAsnGluGlyIleSerAsnLysThrThr	523
QY	988	GATGAGCAAGATGATCGTTCTCTCCAAACAGCAAAAT-----AAACTTCCTCAAGT	1035
Db	524	AspGluAsnAsnAspGluLysIleAspAspAlaAsnTyrArgGluIleAsnGluSerGlu	543
QY	1036	TGCCAAGCAAGAGAGCTGATGGCATCTTTAGGAAACGGTATTTGACTAAGGGAAGTGA	1095
Db	544	SerLeuMetGlnAspLysAlaAsnAspLeuLysPheAsnAspGluValAsnAsnSerGlu	563
QY	1096	GTTAGATTGCATTTCCCAATTTGAGGAGAAAAATATGCTGGGACGACGTACTTAATGCC	1155
Db	564	-----AsnGlnGlnAsnSerSerGluAsnAsnIleAsnAsnAlaValArgAsnAla	580
QY	1156	AAGCCATCTCGA-----AACTCTTCTAGCCTTAATGTAGAGTGTGCAAGT	1200
Db	581	ValSerSerAspIleGluTyrAlaThrAsnGluAspGluAsnAspGlu---ArgLeu	599
QY	1201	TCCAAGCAGCATGGAATAAGGGATCTCAAAAT-----1233	
Db	600	AlaGlnAspThrAsnLysGluAspGlnLysLeuSerGlnSerGluAspIleGlnHisGlu	619
QY	1234	-----1245	
Db	620	SerLeuAsnAsnGluAspValSerLeuThrSerAsnLysThrAspAspSerGluHisLeu	639
QY	1246	ATGACATTTCCCAAGTCAGAGGACAGAAAGAAACAATGTGAAGTCAGACATCAAGA	1305
Db	640	GlulysAspSerLeuAsnGluAspLysLysAlaGlu-----651	
QY	1306	ACAGAAAGGAGATTCCAAATATACATCCACCTTAACCTCCCTCCAGAGAGAAGTGGCTG	1365
Db	652	-----ProSerPheAsnLysThrAsnLysAlaProGlnLysMetSerIle	666
QY	1366	GGA-----ACTCCTATTGAGGAATAGCAAAATG	1395
Db	667	LysProGlySerLysProPheAsnValMetThrProSerAspLysLysArgValMet	686
QY	1396	-----CCTCGGTGGGATCCATTTCCCTTCTTAAGACCACT	1434
Db	687	AspAlaLysLysAsnSerValSerArgAsnLysValAsnValProGluLeuLysProGlu	706
QY	1435	GCAAGTCACACAGTACTGTTGGGTAGACCTTCTGAGAGCAGGAGAGGTTCCGAAACCT	1494
Db	707	ThrLysLysGluAla-----GlnAspGluLysMetAsnAlaGlu	719
QY	1495	TTTCCAACACATTTACAAAGATTTGTGGGATACAAACAT-----1533	
Db	720	PheAspAsnHisLeuAsnGluSerGlnLeuAsnSerAspGluSerSerAspPheAsnVal	739
QY	1534	GTCAAAATGCCTTGTTCGGAAACAAACTGTACCCCTGTGGGAAGATGAG-----AAT	1584
Db	740	AlaSerLeuGluTyrAsnGluSerSerGluHisSerValGluLysAspAsnIleIleAsn	759
QY	1585	GGTGAGCAACTCGAGGAGTAGGTGGGAGCTCATTCAGACTGCACCTTCTC-----1635	
Db	760	AspGluAsnThrArgGluAsnGluHisGlnAspValAspAsnSerGlnAsnAsnAspMet	779
QY	1636	-----AACAAATTCACAGCCCCAGAACTTGAAAGATGCGATTTCTGAAATACAAT	1686
Db	780	ProLysGlyAsnGlnPheSerLysValGlnAsn-----SerAsn	792
QY	1687	GTGGCATATCTCAAGAAATGGCACTTTACAGCTTTGGTT-----GATTTCGGGATAG	1740
Db	793	AsnGlnAsnAspAsnLysHisAspIleAsnGluPheValSerLysLysGluLysThrSerGlu	812
QY	1741	GTACTT-----GAAGAAGCAGAGGCGCCCAACATTTATATCATGTCGCAAT	1782



QY 2302 GTTTCACGGCTGTCACCTAGGCTGCTGGATCAACAATGAGTGTCTTATTATATACAGAGTACT 2361  
Db 216 -----TyrAsnGluArgIleGluIleValGlyVal 225  
QY 2362 GAACAGTACAGTGAATACACAGGCTATGCTGAACCTATATGTTGGGCGCCGAAGCCATGAA 2421  
Db 226 GluArgPheSerGlyThrAspHisAlaSerPheArgPheSerArgAlaLysAla 245  
QY 2422 GATGGGAGTGAAGACGATTGGCAGCGGCTGCACGAGATCGTTGTCATTCAGCGCA 2481  
Db 246 GluArgGluGluAspProValGlyArgGlyThrSerAspLeuSer----- 262  
QY 2482 CTTCACTTCAGACGCTACCTCGATCATGTTGTCCTCCGAGAAAGTG 2526  
Db 263 -----GluLysTyrPheProProlysMet 270

## RESULT 11

US-10-115-482-18  
; Sequence 18, Application US/10115482  
; Publication No. US20030212257A1  
; GENERAL INFORMATION:  
; APPLICANT: Spvtek, et al.  
; TITLE OF INVENTION: NOVEL HUMAN PROTEINS, POLYNUCLEOTIDES ENCODING THEM  
; TITLE OF INVENTION: AND METHODS  
; TITLE OF INVENTION: OF USING THE SAME  
; FILE REFERENCE: 21404-322D  
; CURRENT APPLICATION NUMBER: US/10/115,482  
; CURRENT FILING DATE: 2002-04-05  
; PRIOR APPLICATION NUMBER: 60/281,086  
; PRIOR FILING DATE: 2001-04-03  
; PRIOR APPLICATION NUMBER: 60/281,136  
; PRIOR FILING DATE: 2001-04-03  
; PRIOR APPLICATION NUMBER: 60/281,863  
; PRIOR FILING DATE: 2001-04-05  
; PRIOR APPLICATION NUMBER: 60/281,906  
; PRIOR FILING DATE: 2001-04-05  
; PRIOR APPLICATION NUMBER: 60/282,934  
; PRIOR FILING DATE: 2001-04-10  
; PRIOR APPLICATION NUMBER: 60/283,512  
; PRIOR FILING DATE: 2001-04-12  
; PRIOR APPLICATION NUMBER: 60/285,325  
; PRIOR FILING DATE: 2001-04-19  
; PRIOR APPLICATION NUMBER: 60/285,890  
; PRIOR FILING DATE: 2001-04-23  
; PRIOR APPLICATION NUMBER: 60/286,068  
; PRIOR FILING DATE: 2001-04-24  
; PRIOR APPLICATION NUMBER: 60/286,292  
; PRIOR FILING DATE: 2001-04-25  
; PRIOR APPLICATION NUMBER: 60/287,213  
; PRIOR FILING DATE: 2001-04-27  
; PRIOR APPLICATION NUMBER: 60/288,257  
; PRIOR FILING DATE: 2001-05-02  
; PRIOR APPLICATION NUMBER: 60/291,134  
; PRIOR FILING DATE: 2001-05-15  
; PRIOR APPLICATION NUMBER: 60/282,020  
; PRIOR FILING DATE: 2001-04-06  
; PRIOR APPLICATION NUMBER: 60/291,725  
; PRIOR FILING DATE: 2001-05-17  
; PRIOR APPLICATION NUMBER: 60/294,771  
; PRIOR FILING DATE: 2001-05-31  
; PRIOR APPLICATION NUMBER: 60/296,965  
; PRIOR FILING DATE: 2001-06-08  
; PRIOR APPLICATION NUMBER: 60/299,128  
; PRIOR FILING DATE: 2001-06-08  
; NUMBER OF SEQ ID NOS: 149  
; SEQ ID NO 18  
; LENGTH: 1489  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-115-482-18

Alignment Scores:

Pred. No.: 0.000162 Length: 1489

Score: 172.00 Matches: 161  
Percent Similarity: 33.45% Conservative: 115  
Best Local Similarity: 19.52% Mismatches: 301  
Query Match: 248 Indels: 248  
DB: 15 Gaps: 39  
US-09-302-812-5 (1-3814) x US-10-115-482-18 (1-1489)  
QY 250 AACAAAGAACTATTACTACTTGGATG-----GATACATAAGGACCCACAGACAGCT 300  
Db 418 LysGlnArgValLeuThrAsnTyrMetPheProGlnGlnProArgGlyAspGluAspPhe 437  
QY 301 GAATCAGAAAGTAAAGAAACAAACATACAGAAATGCTCCATCATGATGATTTCTGTGCAG 360  
Db 438 GlnSerAspSerPheAsnProThrLeuTyrGluGluGlnArgGlnGlnArgMet 457  
QY 361 AAAGATAACTTT---TACCCACATAGGTGGAAAAATTTGAAAAATTTCTCTAGCTAAAT 417  
Db 458 ThrValAlaPheGluPheGluAspLysGluAspAspGluAsnAlaGlyLysValLys 477  
QY 418 CTTGAT---AAATCACCACACAGAAAGAGTTTACACAGTATTTCAGAACCAACAGCAG 468  
Db 478 LeuSerCysGlnAlaProTyrGluArgGlyGlnArgGlyLeuThrLeuGlnProAlaArg 497  
QY 469 ---ACTGCGAGTGTGTC---AAGTGGCAGAAATGAAGGAAGCATGCAGAACAGCTT--- 519  
Db 498 LeuSerGlyAspCysThrProTyrAlaArgCysAspGlnGlnIleGlnAspMetPro 517  
QY 520 -----TTGGCAAGTGAGCCT 534  
Db 518 ValProGlnAsnAspProGlnLeuAlaTyrGlyCysIleSerGlyLeuGlnGlnGluArg 537  
QY 535 CCGCGGGGAGCTCCGCTACCAAGAGCAGCTTAGTAATGCT----- 573  
Db 538 SerMetCysThrProLeuProValAlaAlaGlnSerThrLeuProSerLeuSerGly 557  
QY 574 -----AACATTGCTCAGTCACCCACACTGTATGATGATGATGATGATGATGATGATGAT 618  
Db 558 ArgGlnValGluLeuLeuLysArgTyrPro-----ThrPro 570  
QY 619 CATGAAGAGACAGACAGCAATCAGCAGTTTCTTACCTATATAAACTTGCAATACAAAG 678  
Db 571 TyrProGluAspLeuLysAsn-----MetValLysSerValGlnAsnLeuValGlyLys 588  
QY 679 CCAACAGTAGGAGATGGGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 738  
Db 589 ProSerHisGlyValArgValGluAsnSerAsn---ProThrAlaAsnThrGluGlnThr 607  
QY 739 GTGAAAGACTGTACAGGCTGTCAACAGAGGAGGAGGTGGATGCTACACAGAGAGAGCTTTTG 798  
Db 608 ValLysGlu-----LysTyrGluHisLysTyrProValAlaProlLys----- 621  
QY 799 TCAGATGTTGGTGGCGAGGACATTGGAACTGGACCAAAATATGACAAATATGCTGGA 858  
Db 622 ---GluIleThrValGluAspSerPheValHisProAlaAsnGluMetArgIleGlyGlu 640  
QY 859 CAAGAAACAGCCTAGGTGCTCCCTT----- 891  
Db 641 LeuHisProSerLeuAlaGluThrProLeuTyrProProlLysLeuValLeuLeuGlyLys 660  
QY 892 ---GAGAAAGAAAGTACCTGAGTCACTCAATGATGTAGACAACTCCAGAAACAGTTGT 948  
Db 661 AspLysLysGluSerThrAspGluSerGluValAspLysThrHisCysLeuAsnSer 680  
QY 949 CAAGATTCAGAGCAGATGAAGAAACAGTCCAGTCTTTTGATGAGCAAGATGATGCTTC 1008  
Db 681 ValSerSerGlyThrTyrSerAspTyrSerPro-----SerGlnAlaSerSerGlySer 698  
QY 1009 TCCCAACACAGCAATAAATTTCAAGTTGCCAAGCA----- 1044  
Db 699 SerAsnThrArgValLysValGlySerLeuGlnThrThrAlaLysAspAlaValHisAsn 718  
QY 1044 ----- 1044



## RESULT 9

US-10-424-599-156445  
; Sequence 156445, Application US/10424599  
; Publication No. US20040031072A1

## ; GENERAL INFORMATION:

; APPLICANT: La Rosa Thomas J  
; APPLICANT: Kowalic David K  
; APPLICANT: Zhou Yihua  
; APPLICANT: Cao Yongwei  
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
; FILE REFERENCE: 38-21(53223)B  
; CURRENT APPLICATION NUMBER: US/10/424,599  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 285684  
; SEQ ID NO 156445  
; LENGTH: 180  
; TYPE: PRT  
; ORGANISM: Glycine max  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT3847\_112291C.1.pep  
US-10-424-599-156445

## Alignment Scores:

Pred. No.:	3,69e-05	Length:	180
Score:	175.00	Matches:	47
Percent Similarity:	42.94%	Conservative:	23
Best Local Similarity:	28.83%	Mismatches:	43
Query Match:	2.58%	Indels:	50
DB:	12	Gaps:	4

US-09-302-812-5 (1-3814) x US-10-424-599-156445 (1-180)

QY	2533	GAGCTTACAGGCTTACTCGGATTCCTC-----CGTCC-----	2568
Db	7	GlulieAsnLysAlaPheCysGlyPheLeuTyrGlnCysLysTyrGlnProTyrGlnLys	26
QY	2568	-----	2568
Db	27	IleLeuGlnGluAsnGlyCysThrSerAlaLeuPheTyrAlaAlaThrSerThrSerMet	46
QY	2568	-----	2568
Db	47	GlutThrAspGluGlyGluIleSerAsnHisLysIleThrAsnSerGlnAsnAspTyrHis	66
QY	2569	GGAGTTCCTCTGAAATCTTCTGAGTGGCCAGGAACTGGGGCTGGGTGGCTTT	2628
Db	67	GlyMetAspGlnGlyAsnAsnIleGlyValAlaThrGlyAsnTrpGlyCysGlyAlaPhe	86
QY	2629	GGGGGTGACGCTAGATTAAGCCCTTAATACAGATCCTGGCAGTCTGGCGGTGAACGT	2688
Db	87	GlyGlyAspProGluValLysThrIleIleGlnTrpLeuAlaIleSerGlnAlaLeuArg	106
QY	2689	GAC---GTGGTTATTTTCACCTTTGGGACTCAGATTGATGAGACATTTACAGCATG	2745
Db	107	ProPheIleAlaTyrTyrPheGly---LeuGluAlaLeuSerLeuAspGluVal	125
QY	2746	CACACTTCTCTACCGAGGAGCAAGCTGATGTTGGAAAAGTGATACAGTTATGCTTAGA	2805
Db	126	AlaHisTrpIleLeuSerGlnArgTrpThrValGlyAspLeuTrpAsnMetLeuIleGlu	145
QY	2806	TACTACATGAAGATGAGAAATCTGTTCCACCCCTGGACAGACATCAAGCTTTATCCA	2865
Db	146	TyrSerIleAsnArgSerLysGlyGluThrAsnValGlyPheLeuGlnTrpLeuLeuPro	165
QY	2866	TTCATATAC	2874
Db	166	SerIleTyr	168

## RESULT 10

US-10-424-599-233915  
; Sequence 233915, Application US/10424599  
; Publication No. US20040031072A1

## ; GENERAL INFORMATION:

; APPLICANT: La Rosa Thomas J  
; APPLICANT: Kowalic David K  
; APPLICANT: Zhou Yihua  
; APPLICANT: Cao Yongwei  
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
; FILE REFERENCE: 38-21(53223)B  
; CURRENT APPLICATION NUMBER: US/10/424,599  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 285684  
; SEQ ID NO 233915  
; LENGTH: 300  
; TYPE: PRT  
; ORGANISM: Glycine max  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT3847\_5324C.1.pep  
US-10-424-599-233915

## Alignment Scores:

Pred. No.:	7,34e-05	Length:	300
Score:	172.50	Matches:	55
Percent Similarity:	34.58%	Conservative:	47
Best Local Similarity:	18.64%	Mismatches:	64
Query Match:	2.55%	Indels:	129
DB:	12	Gaps:	5

US-09-302-812-5 (1-3814) x US-10-424-599-233915 (1-300)

QY	1729	TTCTGGGATTAAGTACTTGAAGAGCAGAGGCCCAACATTATATCATCATTCATTTCCT	1788
Db	76	PhePheAspGluValMetSerGlyGluGluSerLysTrpPheGlnGluValLeuPro	95
QY	1789	GACATGTGAAATTCACATCTGCTGCCAAATATT-----	1824
Db	96	AlaLeuGlyAsnLeuLeuLeuArgLeuProSerLeuLeuGluSerHisTyrGlnAsnThr	115
QY	1825	-----TGACCCAGCCCAATACCACTCTCTG	1848
Db	116	AspAsnMetAlaIleAspGlyGluAlaGlyAlaMetLeuThrAlaLeuArgLeuLeu	135
QY	1849	AAACAGAAAGATGAATCATCTGTGCATGTCACAGGAACAGATCGGCAGCTTTTACGT	1908
Db	136	AspSerGlnGlnProGlyIleValPheLeuThrGlnGluLeuIleAlaLeuLeuSer	155
QY	1909	AATGCTTTCTTCGACATTTCCCGACGGAATGCCAAGATGAATCGGAGTATTTCTAGT	1968
Db	156	CysSerLeuPheCysLeuPhePro-----ValSerAsp	166
QY	1969	TACCCA-----GACATTAACCTTCAATCGTTTGTGAA	2001
Db	167	ArgProValIleHisLeuProMetIleAsnPheAspValLeuPheGlySerLeuTyrAsp	186
QY	2002	GGAGTTCATCAGGAAACAGAAACTGAAACACTCTCTGCTACTTTCGAGAGATC	2061
Db	187	AspTyrSerGlnLysGlnGluAsnLysIleTrpCysIleValHisTyrPheGlnArgIle	206
QY	2062	ACAGAGAAAAAACCCTCAGGATTTGTCATTTACAGACAGAGCTCTTGAAGATTTTCCA	2121
Db	207	SerSerGluMetProLysGlyIleAla-----	215
QY	2122	GAATGGAAAGGTGTGAAAGACCTTCGACGCTTACACGCTTACAGAGGTACCATTA	2181
Db	215	-----	215
QY	2182	GAAGGAAACGGCCGAGGCATGCTACAGTGGATTTTGCATAATCGTTTGTGGAGGTGT	2241
Db	215	-----	215
QY	2242	GTGACTGGTCGGGACTTGTACAAGAAGAAATCAGATTTTTTAATCAATCTCGAATTGATT	2301
Db	215	-----	215

Db 1589 SerProThrLysLysSerGluAenGluValLysSerProThrLysLysGluLysSerPro 1608  
QY 553 -----CCAAAGCAGCTTACTATGCTAACTGGT 582  
Db 1609 GluLysSerValValGluGluLysSerProLysGluLysSerProGluLysAlaAsp 1628  
QY 583 CAGTCAACCCACACT -----GAT 600  
Db 1629 AsplysProLysSerProThrLysLysGluLysSerProGluLysSerAlaThrGluAsp 1648  
QY 601 GACCACAGTCACAGATCATGAGAA-----GACAGACACATCAGCAGTTCTTAC 654  
Db 1649 ValLysSerProThrLysLysGluLysSerProGluLysValGluGluLysProThrSer 1668  
QY 655 CCTATAAACTTGCATAAATACAAAGCCAAACAGTAGGAGATGGCAGGCGCAAGCAACTGT 714  
Db 1669 ProThrLysLysGluSerSerProThrLysLysThrAspGluValLysSerProThr 1688  
QY 715 AAGTCAGTGAGATCTGCCAGCTGTGAAAGACTGTACAGGCTGTCAACAGGAGGAGTG 774  
Db 1689 LysLysGluLysSerProGlnThrValGlu----- 1698  
QY 775 GATGTCTACAGAGATCTTTGTCTCAGATGTTGGTCCGAGACATGTGAACTGGACCA 834  
Db 1699 -----GluLysProAlaSerProThrLysLysGlu----- 1708  
QY 835 AAAAATGACAACTTACTGCGCAAGAA-----AGCAGCCTAGTGATTCGCTCCATTT 891  
Db 1709 LysSerProGluLysSerValValGluGluValLysSerProLysGluLysSerProGlu 1728  
QY 892 GAGAAGAAGTAGGCTGAGTCACCAATGGATGTAGACAACTCGAAGAACAGTGTCTCA 951  
Db 1729 LysAlaGluLysProLysSerPro-----ThrLysLysGluLysSerProGlu 1745  
QY 952 GATTGAGAAGCATCAAGAAACAGTCCAGTCTTTGATGACGACAGATGATCTCTCC 1011  
Db 1746 LysSerAlaAlaGluLysSerProThrLysLysGluLysSerProGluLysSer 1765  
QY 1012 -----CAAAACAGCAATAAACTTCAAGTTGCCAAGCAAGAGAGCTGAT 1056  
Db 1766 AlaGluLysProLysSerProThrLysLysGluSerSerProValLysMetAlaAsp 1785  
QY 1057 GCGCATCTTAGGAACGGTATTGACTAAGGAGTGAAGTAGATGATGATTTCCAAATTT 1116  
Db 1786 AspGluValLysSerProThrLysLysGluLysSerProGluLysValGlu----- 1802  
QY 1117 GAAGGAGAAATAAGTCTGGACCGAGTAAATGCCAGCCATCTGGAACCTTTCT 1176  
Db 1803 -----GluLysProAlaSerProThrLysLysGluLysSerProGluLysSerAlaAla 1820  
QY 1177 AGCCTTAATGTAGAGTGAAGTTCCAAAGCAGCATGGAAGGAGGATTTCTAAATTA 1236  
Db 1821 Glu-----GluLysSerProThrLysLysGluLysSerProSerProThr 1837  
QY 1237 GATCATTTCAGATTTCCAGTCAGGAGCAGAGAGAAAGAACATGT---GAAGTC 1293  
Db 1838 -----LysLysThrGlyAspGluSerLysGluLysSerProGluLys 1851  
QY 1294 AGACATCAAGAACAGAAAGAGATTTCCAAATATACATCCACCTAACTCCCTCCAG 1353  
Db 1852 ProGluGluLysProLysSerProThrProLysLysSerProGluLysSerProLys 1871  
QY 1354 AAGAGTGGCTGGGAACCTCTATTGAGGAATGAGAAATGCCTGGTGGGATCCAT 1413  
Db 1872 LysLys-----SerLysSerProGluAlaGluLysProPro----- 1883  
QY 1414 TTGCCCTTCTTAAGACCATCTCCAGTCACACAGTGCATGTTGGGTAGACCTTCGAGA 1473  
Db 1884 AlaProLysLeuThrArgAspLeuLysLeuGlnThrValAsnLysThrAspLeuAlaHis 1903  
QY 1474 GCAGGAGAGTTCCGAAACCTTTCCACACATTAACAGATTTGTGGATACAAACAT 1533  
Db 1904 Phe-----GluValValGluHisAlaThrGluCysLysLysTrpPheLeuAspGlyLysGlu 1922

QY 1534 GTGAATAAGCTTGTTCGGAACAAACTTGTACCTGTGCAAGATGAGAATGGTAGCGA 1593  
Db 1923 Ile-----ThrThrAlaGlnGlyValThrValSerLysAspAspGlnPheGluPhe 1939  
QY 1594 ACTGCGAGGAGTAGGTGGGAGCTTATTCAGACTGCACCTTCTCAACAAATTCACACGACC 1653  
Db 1940 ArgCysSer-----IleAspThrThrMetPheGlySerGlyThr----- 1952  
QY 1654 CAGAACTTGAAGATCGGATTCGAAATACAAATGATGTCATATCT----- 1998  
Db 1953 -----ValSerValValAlaSerAsnAlaAlaGlySerValGluThrLysThr 1968  
QY 1699 -----AAGAAATGGGACTTTTACA----- 1716  
Db 1969 GluLeuLysValLeuGluThrProLysGluThrLysLysProGluPheThrAspLysLeu 1988  
QY 1717 -----GCTTTGGTGAATTTCTGGATAAGGTACTTGAAGAAGCAGAGGCCCAACAT--- 1767  
Db 1989 ArgAspMetGluValThrLysGlyAspThrValGlnMetAspValIleAlaLeuHisSer 2008  
QY 1768 ---TTATATCAG-----TCC 1779  
Db 2009 ProLeuThrLysTrpTrpGlnAsnGlyAsnLeuLeuGluAspGlyLysAsnGlyValThr 2028  
QY 1780 ATTTTACCTGACATGTTGAAATTTGCACCTGTCTGCCAAATATTTGCACCCAGCCAA 1839  
Db 2029 IleLysAsnGluGluAsnLysSerSerLeuIleIleProAsn-----AlaGlnAspSer 2046  
QY 1840 CCATCTCTGAAACAGAAAGATGAATCTCTGTCACGATGTCAAGGACAGATCGCCAGT 1899  
Db 2047 GlyLysIleThrValGluAlaSerAsnGluValGly---SerSerGluSerSerAlaGln 2065  
QY 1900 CTTTGTAGCTAAGCTTCTCTGCGACA-----TTTCCCGCA----- 1935  
Db 2066 LeuThrValAsnProProSerThrThrProIleValValAspGlyProLysSerValThr 2085  
QY 1936 -----CGGAATGCCAAGATGAATCGGAGTATTCGAGTATTTAGTTACCCA----- 1974  
Db 2086 IleLysGluThrGluThrAlaGluPheLysAlaThrIleSerGlyPheProAlaProThr 2105  
QY 1975 -----GACATTAACCTCAATCGTGTGTTGAAGGAGCTTCATCAAGGAAACAGAA 2025  
Db 2106 ValLysTrpThrIleAsn---GluLysIleValGlu-----GluSerArgThrIleThr 2122  
QY 2025 AAATGAAACACTCTCTGCTACTTTTCGAGAGTCACAGAGAAATA-----CCTACA 2079  
Db 2123 ThrIleLysThrGluAspValThrThrLeuLysIleSerAsnAlaLysIleGluGlnThr 2142  
QY 2080 GGATTTGGTGACATTTTCAAGACAGAGT-----CTTGAAGATTTT 2118  
Db 2143 GlyThrValLysValThrAlaGlnAsnSerAlaGlyGlnAspSerLysGlnAlaAspLeu 2162  
QY 2119 CCAGATGGGAAAGGTGTAAGCTCTGACACGCTTACACGCTTACAGGTACCTACAGGGTACC 2178  
Db 2163 LysValGluProAsnValLysAlaProLysPheLysSerGlnLeuThrAspLysValAla 2182  
QY 2179 ATAGAGGCACACGGC---CGAGGCATGCTACAGGTGGATTTTGCAGATCGTTTGTGGA 2235  
Db 2183 AspGluGlyGluProLeuArgTrpAsnLeuGluLeuAsp----- 2195  
QY 2236 GGTGTGTGACTGGTGGCGGACTTGTACAGAGAAATCAGATTTTATCAATCCTGAA 2295  
Db 2196 GlyProSerProGlyThr-----GluValSerTrpLeuAsnGlyGln 2210  
QY 2286 TTGATTTTTCACGCTGTTCACTGAGGTGCTGGATCACAATGAG---TGCTTTATTATC 2352  
Db 2211 ProLeuThrLysSerAspThrValGlnValValAspHisGlyAspGlyThrTyrHisVal 2230  
QY 2353 ACAGGTACTGACAGTACAGTGAATACACAGC 2385  
Db 2231 ThrIleAlaGluAlaLysProGluMetSerGly 2241

QY 2782 AAA 2784  
Db 719 LVS 719

## RESULT 7

```

RES001
US-10-424-599-184988
; Sequence 184988, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Mole
; TITLE OF INVENTION: Plants and Uses Ther
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 184988
; LENGTH: 200
; TYPE: PR
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(200)
; OTHER INFORMATION: unsure at all Xaa loc
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT MRT3847
US-10-424-599-184988

```

Alignment Scores:	
Pred. No.:	8,19e-09
Pred. Score:	220.50
Percent Similarity:	45.20%
Best Local Similarity:	30.51%
Query Match:	3.26%
DB:	12
Length:	200
Matches:	54
Conservative:	26
Mismatches:	68
Indels:	29
Gaps:	4

US-09-302-812-5 (1-3814) X US-10-424-599-184988 (1-200)

1876	QY	ATGTCACAGAA	CAGATGCGCAGTCTTTT	TAGCTAATGCTTTTCTCTGCA	CAATTTCCCGA	1935
33	Db	LeuThrGlnGlu	ThrSerAlaProProSerArg	ThrLeuPheCysLeuPhePro	---	51
1936	QY	CGGAATGCCAAG	ATAATCGGATGATTCTAGTTACCA	---	---	1974
52	Db	-----	-----	ValSerAspArgProGlyIleHisLeuProMetIle	63	
1975	QY	-----	GACATTAACTCAATCGGTTGTTGAAGCAGCTTCATCAAGAAACCGAAAAA	2028		
64	Db	AsnSerAspValSerSerGlySerLeuTyrGluAspTyrSerGlnLysProGluSerThr	83			
2029	QY	CTGAAAAACACTCTCTGCTACTTCTCGAAGAGTCA	CAGAGAAAAACCTACAGAGTTCGGT	2088		
84	Db	ThrGlyArgIleAlaHisTyrGlyGlnArgIleSerSerGluMetProLysGlyIleVal	103			
2089	QY	ACATTACAAGCAGCAGTCTTGAA	-----	-----	GATTTTCCAGAA	2124
104	Db	SerPheGluArgLysValLeuProPheLysAsnAspSerIleHisIleSerTyrProAsp	123			
2125	QY	-----	TGGAAAGGTGTGAAAACCTCTGACACGCTTACACGTCACCTTACAGGGT	2175		
124	Db	AlaAsnPheTrpSerThrSerAlaIleProLeuCysArgPheGluValHisSerSerGly	143			
2176	QY	ACCATGAAGGCACCGCGCAGGCATGCTCAGGTGATTTTGCAATCGTTTGTTCGA	2235			
144	Db	LeuIleGluAspGlnSerSerGlyAlaAlaGluValAspLeuAlaAsnLysTyrLeuGly	163			
2236	QY	GGTGGTGTGACTGGTCGGGACTTGTACAGAGAAATCAGATTTTAAATCAATCTGAA	2295			
164	Db	GlyGlyAlaLeuGlyArgGlyCysValGlnGluGluIleArgPheMetValSerProGlu	183			

Qy	2296	TTGATTGTTT	CACGGCTGTT	CACTGAGGTG	CTGGATCACA	ATGAGTGTCTT	2346
				::			
Dd	184	LeuGluAla	GlyMetLeu	PheLeuPro	AlaMetAla	AspAsnGluAla	200

## RESULT 8

```

US-10-369-493-5013
; Sequence 5013, Application US/10369493
; Publication NO. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slatler, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Cher, Xiaofeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: 2003-02-28
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 5013
; LENGTH: 6642
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
US-10-369-493-5013

Alignment Scores:
Pred. No.: 0.000147 Length: 6642
Score: 176.00 Matches: 187
Percent Similarity: 34.46% Conservative: 120
Best Local Similarity: 20.99% Mismatches: 372
Query Match: 2.60% Indels: 212
DB: 15 Gaps: 42

US-09-302-812-5 (1-3814) X US-10-369-493-5013 (1-6642)

```

88	QY	TTCTGCGCCGCACTCCCTCTGGAC---TCCGGAGAGCTTCCTT-----GGCAGGCAG	132
1438	Db	SerSerProThrLysLysGluLysSerProSerProThrLysLysThrGluAspGlu	1457
133	QY	AGCGGTGTTCTGACCCCAAGGACGCTCCCGTCCAGTTCCAGGTCCCTCCGTCCTCGCCCA	192
1458	Db	ValLysGluLysSerProProLysSerProThrLysLysGluLysSerProGluLysPro	1477
193	QY	GCCTGCTGCTCGGGGGCGGACCGCACAGAGGC---AACGCCACCTCGTTTGT---	246
1478	Db	GluAspValLysSerProValLysLysGluLysSerProAspAlaThrAsnIleValGlu	1497
247	QY	-----TTCAACAACAAAGACTATTACTACTTCGGATGGATACT	282
1498	Db	ValSerSerGluThrThrIleGluLysThrGluThrThrMetThrThrHis	1517
283	QY	AAAGGACCCCAAGACACGCTGAATCAGAAAGTAAAGAAAAACAACAATCAAGAAATTGACTCC	342
1518	Db	GluSerGluGluSerArgThrSerValLysLysGluLysThrProGluLysValAspGlu	1537
343	QY	ATCATGAGTTCGTGCGAAGAAGATACTTTTCCACATAAG-----GTGGAATAATTG	396
1538	Db	LysProLysSerProThrLysLysAspLysSerProGluLysSerIleThrGluGluIle	1557
397	QY	GAATAATGTTCTCAGCTAATCTTGATAAATCACCACA-----GAAAAGATTCCACAG	450
1558	Db	-----LysSerProValLysLysGluLysSerProGlu	1568
451	QY	TATTTGAACCAACAGCAGACTCGAGTGTGTGCAAGTGGCAGAAATCAAGGGAAGCATGCA	510
1569	Db	LysValGluGluLysProAlaSerProThrLysLysGluLysSerProGluLysProAla	1588
511	QY	GAACAGCTTTTGGCAAGTAGGCTCCCGGGGAGACTCCGCTA-----	552



```

Db      108 AlaAlaAspGlyLeuAlaLeu-----PhePheAspAspLeuLeuSerArgAlaGlnAla 125
QY      1762 CAACATTTATATACAGTCATTTTACCTGACATGGTGAATAATGCACTGTCTGTGCCAAT 1821
Db      126 ArgGlyTrpPheSerGluValProAenLeuAlaAargLeuLeuLeuLeuArgLeuProThr 145
QY      1822 ATTTGCAACCCCAATACCACTCTCTGAAACAGAGATCAATCATCTCT-----1869
Db      146 LeuLeuGluAspHisTyr-----AlaLysAlaGlyHisGlyAlaSerGlyLeu 161
QY      1870 -----GTCCAGATGTCACAGAAACAGATCGCCAGT 1899
Db      162 ArgValMetAlaSerGlnAspAlaGlyValValLeuLeuLeuSerGlnGluLeuValAlaAla 181
QY      1900 CTTTGTAGCTAATGCTTCTCTGCACTTTCCCGACGGAATGCCAGATGAATCGGGAG 1959
Db      182 LeuLeuThrCysAlaLeuPheCysLeuPheProThrAlaGlyArgAlaGln-----198
QY      1960 TATTCTAGTTACCCAGACATTAATCACTCGTTGTTT-----GAAGGA 2004
Db      199 ---AlaCysLeuProThrIleAsnPheAspGlyLeuPheAlaProLeuIleHisAsnAla 217
QY      2005 CGTTCATCAAGAAACCCAGAAACCTGAAACACTCTTCTGCTACTTCTCGAAGACTCACA 2064
Db      218 ArgGlnSerGlnGlu---GlnLysValArgCysLeuValHisTyrPheGluArgValThr 236
QY      2065 GAGAAAAACCTACAGGATGCTGACATTTCAACAGACAGATCTT-----2109
Db      237 AspSerMetProThrGlyLeuValSerPheGluArgLysValLeuProArgAlaLeu 256
QY      2110 -----GAAGATTTCCAGATTCGGAAGAGGTGTGAAAGCCTCTGACA 2151
Db      257 SerAspGlyValProTyrProAspIleHisAlaTrpValAlaSerSerAlaProLeuCys 276
QY      2152 CGCTTACACGTCACATTACAGGGTACCATAGAAGCAACGCCCGAGGATGTCACAGGTG 2211
Db      277 GlnPheThrValPheSerSerGlyPheIleGluAspGluGluGlnGluAlaLeuGlnVal 296
QY      2212 GATTTTCAATCGTTTGTGGAGTGTGTGTGCTGCTGCTGCGGACCTGTGTACAGAA 2271
Db      297 AspPheAlaAsnLysTyrLeuGlyGlyAlaLeuSerTrpGlyCysValGlnGluGlu 316
QY      2272 ATCAGATTTTAAATCAATCTGATTTGATTTTTCACGGCTGTTCACTGAGTGTGATG 2331
Db      317 IleArgPheMetIleAsnProGluLeuIleValGlyMetLeuPheLeuSerCysMetGlu 336
QY      2332 CACATGAGTGTCTTATATACAGGTACTGAACAGTACAGTGAATACACAGGCTATGCT 2391
Db      337 AspAsnGluAlaIleGluIlePheGlyAlaGluArgPheSerGlnTyrMetGlyTyrGly 356
QY      2392 GAAACTTATCGTTGGCCCGAAGCCATGACATGGAGTGAAGAGGAGGATGGCAGCGG 2451
Db      357 SerSerPheArgPheValGlyAspTyrLeuAspThrLysProPheAspSerMetGlyArg 376
QY      2452 CGCTGACGAGATCGTTGTCATTCAGCCACTTCACCTTCAACGCTACCTCGATCAGTTT 2511
Db      377 ArgArgThrArgIleValAlaIleAspAlaLeuAspCysProAlaArgLeu---HisTyr 395
QY      2512 GTCCCTCAGAAAGTACAGCGTACCTTAAACAGGCTTACGCGGATTCCTCCGT-----2565
Db      396 GluSerGlyCysLeuLeuArgGluValAsnLysAlaPheCysGlyPheAspGlnSer 415
QY      2566 -----CCTGGA 2571
Db      416 LysHisIleLeuTyrAlaLysLeuPheGlnAspLeuHisAsnLysAspPheSerSer 435
QY      2572 GTTCCTTCTGAAATCTTCTGACAGTCCCGCAAGAACTGGGCTGTGTGCTTTGGG 2631
Db      436 IleAsnSerSerGluTyrValGlyValSerThrGlyAsnTrpGlyCysGlyAlaPheGly 455
QY      2632 GTTCAGCTAGATTAAGGCTTAATACATCTCGCAGCTGTGCGGCTGAACGTGAC 2691

```

```

Db      456 GlyAsnProGluIleLysSerMetIleGlnTrpIleAlaAlaSerGlnAlaLeuArgPro 475
QY      2692 GTGGTT---TATTTCCACCTTTGGGCACTCAGAGTTGATGAGACATTTACAGCATGCAC 2748
Db      476 PheValAsnTyrTyrThrPheGluAspValSerLeuGlnArg---LeuGluGluValIle 494
QY      2749 ACTTTCCTTACGAGAGGAAGCTGGATGTTGAAAGGTACAGTTATGCTTATAGATAC 2808
Db      495 GlnTrpIleArgLeuHisGlyTrpThrValGlyGluLeuTrpHisMetLeuMetGluTyr 514
QY      2809 TACAATGAA 2817
Db      515 SerSerGln 517

RESULT 6
US-09-973-451-10
; Sequence 10, Application US/09973451
; Patent No. US20020132328A1
; GENERAL INFORMATION:
; APPLICANT: JACOBSON, Myron K.
; APPLICANT: JACOBSON, Elaine L.
; APPLICANT: AM, Jean-Christophe
; APPLICANT: LIN, Winston
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE
; TITLE OF INVENTION: (PARG) ENZYMES
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV
; TITLE OF INVENTION: THEREWITH
; FILE REFERENCE: NTAD 201
; CURRENT APPLICATION NUMBER: US/09/973,451
; CURRENT FILING DATE: 2001-10-09
; PRIOR APPLICATION NUMBER: US/09/302,812
; PRIOR FILING DATE: 1999-04-30
; PRIOR APPLICATION NUMBER: 60/083,768
; PRIOR FILING DATE: 1998-05-01
; NUMBER OF SEQ ID NOS: 38
; SEQ ID NO 10
; LENGTH: 726
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
; FEATURE:
US-09-973-451-10

Alignment Scores:
Pred No.: 1,42e-26 Length: 726
Score: 443.50 Matches: 184
Percent Similarity: 36.62% Conservative: 124
Best Local Similarity: 21.88% Mismatches: 278
Query Match: 6.55% Indels: 255
DB: 9 Gaps: 32

US-09-302-812-5 (1-3814) x US-09-973-451-10 (1-726)
QY      622 GAAGAAGACAGACACATCAGACAGTTCCTTACACCTATAAACTTGCAATAACAAAGCCA 681
Db      14 GlnAspGluLysAspTyrGluAspTyrVal---GlyValGlyPheAlaHisGlnValPro 32
QY      682 ACAGTAGGACATGGCGAGCCAGACCACTGAAGTCAGTCAGTGCATCTCCAGTCTGTG 741
Db      33 ThrMet-----LysArgArgLysLeuThrGluHisGlyAsnThrThrGluSerLys 49
QY      742 AAAGACTGTACAGGCTGTCAACAGGAGGAGGTGTGATGTGTACACAGAGTCTCTTTGTCA 801
Db      50 GluAspProGlu-----GluProLysSerArgAspValPheValSerSerGlnSer 67
QY      802 GATGTTGTGTCGCGAGGACATTTGCACTGGACCAAAA-----837
Db      68 AspGluSerGlnGluAspSerAlaGluAsnProGluIleAlaLysGluValSerGluAsn 87
QY      838 -----AATGACAACAATTTG 852
Db      88 CysGluAsnLeuThrGluThrLeuLysIleSerAsnIleGluSerLeuAspAsnValThr 107
QY      853 ACTGGACAAGAGCAGCTAGTGTAT-----TCGCTCCATTGAGAAAGAAAGT 903

```

94 LeuProValThrAlaGlyAsnLeuHisArgValMetTyrGlnLeuProIleArgGlu 113  
QY 1480 GAGGTTCCGAAACCTTTTCCACACATTCACAAAGATTGTGGGATACAAACATGTGAAA 1539  
Db 114 ThrProArgProTyrLysSerProGlyLys-----TrrpAspSerGluHisValArg 131  
QY 1540 ATGCTTTGTTGGGACAAACCTTCACTCCCTGGAAGATGAGAATGGTGGAGCGAATGCA 1599  
Db 132 LeuProCysAlaProGluSerLysTyrProArgGluAsnProAspGlySerThrIle 151  
QY 1600 GGGAGTAGGTGGGAGCTCATTCAGACTGCACTTCTCAACAAATTCACAGACCCAGAAC 1659  
Db 152 AspPheArgTrrpGluMetIleGluArgAlaLeuLeuGlnProIleLysThrCysGluGlu 171  
QY 1660 TTCAAGGATCGGATCTCTGAATCAATGCGCATTCATTAAGAAATGGGACTTTACAGCT 1719  
Db 172 LeuGlnAlaAlaIleIleSerTyrAsnThrThrTyrArgAspGlnTrpHisPheArgAla 191  
QY 1720 TTGGTTGATTCTGGGATAAGTACTTGAAGACGAGCGCCCAACATTTATATCATCTCC 1779  
Db 192 LeuHisGlnLeuLeuAspGluGluLeuAspGluSerGluThrArgValPhePheGluAsp 211  
QY 1780 ATTTTACTGACATGGTGAATATGCACTCTGTCTGCAAAATATTTCACCCAGCCCAATA 1839  
Db 212 LeuLeuProArgIleIleArgLeuAlaLeuArgLeuProAspLeuLeuGlnSerProVal 231  
QY 1840 CCACTCTGAAACAGAGATGATCATCTCTGTCAGCATGTACAGGACAGACAGATCGCCAGT 1899  
Db 232 ProLeuLeuLysHisHisLysAsnAlaSerLeuSerLeuSerGlnGlnIleSerCys 251  
QY 1900 CTTTGTAGTAACTCTTCTGCACTTCTCCGACGATTCCTCCGACGGAATGCC---AAGATGAATCG 1956  
Db 252 LeuLeuAlaAsnAlaPheLeuLysThrPheProArgArgAsnThrLeuLysArgLysSer 271  
QY 1957 GAGTATCTAGTTACCCAGACATTAATCTCAATCGGTGTGTTGAAGACGCTTCATCAAGG 2016  
Db 272 GluTyrSerThrPheProAspIleAsnPheAsnArgLeuTyrGlnSerThrGlyProAla 291  
QY 2017 AAACACAGAAACCTGAAACACTCTTCTGCTACTTTCGAGAGCTC-----ACAGAGAAA 2070  
Db 292 ValLeuGluLysLeuLysCysIleMetHisTyrPheArgArgValCysProThrGluArg 311  
QY 2071 AAA-----CCTACAGGATTTGTCACATTTTACAGACAGAGT-----CTTGAA 2112  
Db 312 AspAlaSerAsnValProThrGlyValValThrPheValArgSerGlyLeuProGlu 331  
QY 2113 GATTTTCCAGATGGAAAGGTGTGAAAGCCTCTG-----ACAGCTTACAGTCACT 2166  
Db 332 HisLeuLeuAspTrpSerGlnSerAlaAlaProLeuGlyAspValProLeuHisValAsp 351  
QY 2167 TACGAGGTACCATAGAGGCAACGCGGAGCATGTCTACAGTGGATTTTCAATCGT 2226  
Db 352 AlaGluGlyThrIleGluAspGluGlyIleGlyLeuLeuGlnValAspPheAlaAsnLys 371  
QY 2227 TTTGTTGAGGTGTGTGACTGTGCGGAGCTGTGTACAGAGAAATCAGATTTTAAATC 2286  
Db 372 TyrLeuGlyGlyValLeuGlyHisGlyCysValGlnGluIleArgPheValIle 391  
QY 2287 AATCCTGAATGATTTTCAAGGCTGTCTACAGTGTCTGATGATCAATGATGTCTT 2346  
Db 392 CysProGluLeuLeuValGlyLysLeuPheThrGluCysLeuArgProPheGluAlaLeu 411  
QY 2347 ATTATCACAGGTACTGAAACAGTACAGTGAATACACAGGCTATGCTGAAACTTATCTGG 2406  
Db 412 ValMetLeuGlyAlaGluArgTyrSerAsnTyrThrGlyTyrAlaGlySerPheGluTrp 431  
QY 2407 GCGCGAAGCCATGAAGATGGGAGTGAAGAGGACGATTTGGCAGCGCGCTGCGAGGAGATC 2466  
Db 432 SerGlyAsnPheGluAspSerThrProArgAspSerSerGlyArgArgGlnThrAlaIle 451  
QY 2467 GTTGCATTTGAGCACTTCACTTACAGCGCTACCTCGATCATGTTGTGCTGAGAAAGTG 2526  
Db 452 ValAlaIleAspAlaLeuHisPheAlaGlnSerHisHisGlnTyrArgGluAspLeuMet 471

QY 2527 AGAGTGAGCTTAAACAGGCTTACTCGGAGTTCCTCCGT-----CCT 2568  
Db 472 GluArgGluLeuAsnLysAlaTyrIleGlyPheValHisTrpMetValThrProPro 491  
QY 2569 GGAGTTCCTTCTGAAATCTTTCTGCAAGTGGCCACGGGAACTGGGCTGTGGTGCCTTT 2628  
Db 492 Gly-----ValAlaThrGlyAsnTrpGlyCysGlyAlaPhe 503  
QY 2629 GGGGGTCACTAGATTAAAGCTTAAATACAGATCTGGCAGCTGCTGGGCTGAACGT 2688  
Db 504 GlyGlyAspSerTyrLeuLysAlaLeuLeuGlnLeuMetValCysAlaGlnLeuGlyArg 523  
QY 2689 GAGTGGTTATTTCACCTTTGGGACTCAGAGTTGATGAGACACATTTACAGCATCCAC 2748  
Db 524 ProLeuAlaTyrTyrThrPheGlyAsnValGluPheArgAspAspPheHisGluMetTrp 543  
QY 2749 ACTTTCCTTACCAGAGGAGCTGGATGTTGGAAAGTGTACAAGTTATTGCTTAGATAC 2808  
Db 544 LeuLeuPheArgAsnAspGlyThrThrValGlnGlnLeuTrpSer---IleLeuArgSer 562  
QY 2809 TACAATCAAGAAATGCAGAAACTGTTCACCCCTGGACCA-----GAC 2850  
Db 563 TyrSerArgLeuIleLysGluLysSerSerLysGluProArgGluAsnLysAlaSerLys 582  
QY 2851 ATCAAGCTTTATCCATTCATATACCATGCTGTGTGAGTCAAGTGCAGAGACCATGACATG 2910  
Db 583 LysLysLeuTyrAspPheIle-----LysGluGluLeuLysLysValArgAspVal 599  
QY 2911 CCAGAGACAGAGGCA 2925  
Db 600 ProGlyGluGlyAla 604

RESULT 5

US-10-425-114-60000  
; Sequence 60000, Application US/10425114  
; Publication No. US20040034888A1  
; GENERAL INFORMATION:  
; APPLICANT: Liu, Jindong  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Screen, Steven E  
; APPLICANT: Tabaska, Jack E  
; APPLICANT: Cao, Yongwei  
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 38-21(53313)B  
; CURRENT APPLICATION NUMBER: US/10/425,114  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 73128  
; SEQ ID NO 60000  
; LENGTH: 546  
; TYPE: PRT  
; ORGANISM: Zea mays  
; FEATURE:  
; OTHER INFORMATION: Clone ID: LIB3689-262-E12\_FLI.pep  
US-10-425-114-60000

Alignment Scores:  
Pred. No.: 1.44e-37 Length: 546  
Score: 579.00 Matches: 150  
Percent Similarity: 49.44% Conservative: 69  
Best Local Similarity: 33.86% Mismatches: 154  
Query Match: 8.55% Indels: 70  
DB: 12 Gaps: 12

US-09-302-812-5 (1-3814) x US-10-425-114-60000 (1-546)

QY 1660 TTGAAGGATCGATTCTGAAATACAAATGTGCAVATTC-----AAG 1701  
Db 88 LeuAlaAspAlaLeuThrAspLeuArgLeuAlaLeuAlaLeuProAlaLeuProArg 107  
QY 1702 AAATGGGACTTACAGCTTGTGTTGTTTCTGGGATAGGTACTTGAAGAGCAGAGGCC 1761

Db 501 ThrHisPheLysAspLeuTrpAspSerHisValLysMetProCysSerGluGlnAsn 520  
Qy 1561 TTGTACCCCTGTGGAGATGAGATGGTGTGAGCGAACTGCAGGGAGTAGGTGGGAGCTCAT 1620  
Db 521 LeuTyProValGluAspGluAsnGlyGluArgAlaIleGlySerArgTrpGluLeuIle 540  
Qy 1621 CAGACTGCCTTCTCAACAATTCACAGCCCGGAGAACTTGAAGAGTGCAGATTCGAAA 1680  
Db 541 GlnThrAlaLeuLeuAsnArgLeuThrArgProGlnAsnLeuLysAspAlaIleLeuLys 560  
Qy 1681 TACATGTGGCATATTTCAAGAAATGGGACTTTACAGCTTGTGGTGTGGATTAAG 1740  
Db 561 TyrAsnValAlaTySerLysTrpAspPheThrAlaLeuIleAspPheTrpAspLys 580  
Qy 1741 GTACTTGAAGACAGAGCGCCCAACATTTATATCATCTAGTCCATTTTACCTGACATGTGAAA 1800  
Db 581 ValLeuGluGluAlaGluAlaGlnHisLeuTyGlnSerIleLeuProAspMetValLys 600  
Qy 1801 ATTGCACCTGTCTGCCAAATATTTGCACCCAGCCCAATACCACTCTCGAAACAGAGATG 1860  
Db 601 IleAlaLeuCysLeuProAsnIleCysThrGlnProIleProLeuLeuLysGlnLysMet 620  
Qy 1861 AATCATTTCTGCACCATGTACAGGAACAGATCGCCAGTCTTTAGCTAATGCTTCTTC 1920  
Db 621 AsnHisSerIleThrMetSerGlnGluGlnIleAlaSerLeuLeuAlaAsnAlaPhePhe 640  
Qy 1921 TGCATTTTCCCGACGGAATGCCAAGATGAATCGAGTATCTAGTTACCCAGACATT 1980  
Db 641 CysThrPheProArgArgAsnAlaLysMetLysSerGluTySerSerTyProAspIle 660  
Qy 1981 AACTTCATCGTGTGTGTGAAGAGCTTCATCAAGGAACACGAAACAACTGAAACATC 2040  
Db 661 AsnPheAsnArgLeuPheGluGlyArgSerArgLysProGluLysLeuLysThrLeu 680  
Qy 2041 TTCTGCTACTTTTCAAGAGTACACAGAAACCTACAGATTGCTGACATTTACAGA 2100  
Db 681 PheCysTyPheArgArgValThrGluLysProThrGlyLeuValThrPheThrArg 700  
Qy 2101 CAGAGTCTTGAAGATTTTCCAGAAATGGGAAGGTGTGAAAGCCCTGCACACGCTTACAC 2160  
Db 701 GlnSerLeuGluAspPheProGluTrpGluArgCysGluLysLeuLeuThrArgLeuHis 720  
Qy 2161 GTCACTTACAGGGTACATAGAGGCAACGGCCGAGCATGCTACAGTGGATTTTGA 2220  
Db 721 ValThrTyGluGlyThrIleGluGlyAsnGlyGlnGlyMetLeuGlnValAspPheAla 740  
Qy 2221 AATCGTTTGTGGAGGTGGTGTGACTGTCGGGACTTGTACAGAGAAATCAGATT 2280  
Db 741 AsnArgPheValGlyGlyValThrSerAlaGlyLeuValGlnGluGluIleArgPhe 760  
Qy 2281 TTAATCAATCTGAATGATTTTTCACGGCTGTTTCACTGAGGTGCTGGATCACAATGAG 2340  
Db 761 LeuIleAsnProGluLeuIleValSerArgLeuPheThrGluValLeuAspHisAsnGlu 780  
Qy 2341 TGTCTTATTATCAGGTACTGACAGTACAGTACAGTAAATACAGAGCTATGCTGAACTTAT 2400  
Db 781 CysLeuIleIleThrGlyThrGluGlnTySerGluTyThrGlyTyAlaGluThrTy 800  
Qy 2401 CGTTGGCCCGCAAGCCATGAAGTGGGAGTGAAGGACGATTTGGCAGCGCGTGCACG 2460  
Db 801 ArgTrpAlaArgSerHisGluAspArgSerGluArgAspPheTrpGlnArgArgThr 820  
Qy 2461 GAGATCGTGGCATGTAGCGACTTCACTTACAGCGCTACCTCGATCAGTTTGTGCTGAG 2520  
Db 821 GluIleValAlaIleAspAlaLeuHisPheArgArgTyLeuAspGlnPheValProGlu 840  
Qy 2521 AAAGTGAGAGCTGAGCTTAACAAGCTTACTCGGATTCCTCCGCTCGGATTCCTTCT 2580  
Db 841 LysIleArgGluLeuAsnLysAlaTyCysGlyPheLeuArgProGlyValSerSer 860  
Qy 2581 GAAATCTTTCTGAGTGGCCACGGGAACTTGGGGCTGTGGTGTCTTTGGGGGTGACGCT 2640  
Db 861 GluAsnLeuSerAlaValAlaThrGlyAsnTrpGlyCysGlyAlaPheGlyGlyAspAla 880

Qy 2641 AGATTAAAGCTTAAATACAGATCTCGCAGCTGCTGGCGCTGAACGTGACGTGTTTAT 2700  
Db 881 ArgLeuLysAlaLeuIleGlnIleLeuAlaIleValAlaGluArgAspValValTy 900  
Qy 2701 TTCACCTTTGGGACTCAGAGTTCATGAGAGACATTTTACAGCATGCACACTTCTTACC 2760  
Db 901 PheThrPheGlyAspSerGluLeuMetArgAspIleTySerMetHisThrPheLeuThr 920  
Qy 2761 GAGAGGAGCTGGATGTGGAAAGTGTACAGTATTGCTTGTAGTACTACATGAAGAA 2820  
Db 921 GluArgLysLeuThrValGlyGluValTyLysLeuLeuArgTyTyAsnGluGlu 940  
Qy 2821 TCCAGAAACTCTTCCACCCCTGGACACACATCAAGCTTTTATCCATTATATACCATGCT 2880  
Db 941 CysArgAsnCysSerThrProGlyProAspIleLysLeuTyProPheIleTyHisAla 960  
Qy 2881 GTTGAGTCAAGTGCAGAGACCACTGACATGCCAGACAGAGAGGCGAGC 2928  
Db 961 ValGluSerCysThrGlnThrThrAsnGlnProGlyGlnArgThrGly 976

RESULT 4  
US-09-973-451-8  
; Sequence 8, Application US/09973451  
; Patent No. US20020132328A1  
; GENERAL INFORMATION:  
; APPLICANT: JACOBSON, Myron K.  
; APPLICANT: JACOBSON, Elaine L.  
; APPLICANT: AM, Jean-Christophe  
; APPLICANT: LIN, Winston  
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE  
; TITLE OF INVENTION: (PARG) ENZYMES,  
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV  
; TITLE OF INVENTION: THEREWITH  
; FILE REFERENCE: NIAID 201  
; CURRENT APPLICATION NUMBER: US/09/973,451  
; CURRENT FILING DATE: 2001-10-09  
; PRIOR APPLICATION NUMBER: US/09/302,812  
; PRIOR FILING DATE: 1999-04-30  
; PRIOR APPLICATION NUMBER: 60/083,768  
; PRIOR FILING DATE: 1998-05-01  
; NUMBER OF SEQ ID NOS: 38  
; SEQ ID NO 8  
; LENGTH: 768  
; TYPE: PRT  
; ORGANISM: Drosophila melanogaster  
; FEATURE:  
US-09-973-451-8

Alignment Scores:  
Pred. No.: 1,65e-77 Length: 768  
Score: 1074.50 Matches: 244  
Percent Similarity: 58.12% Conservative: 96  
Best Local Similarity: 41.71% Mismatches: 194  
Query Match: 15.86% Indels: 51  
DB: 9 Gaps: 15

US-09-302-812-5 (1-3814) x US-09-973-451-8 (1-768)

Qy 1249 AGAATTTCCAGTTCAGAGGACAGAGAAAGAAACAATGTGAAGTCAGACATCAAGAACA 1308  
Db 45 ArgMetSerLysSerProAspGlyGlyIleSerGluIleGluThrGluGluGlu----- 62  
Qy 1309 GAAAGGAAGATTCCAAATACATCCCACTTAACCTCCCTCCA-----GAGAAGAGTGG 1362  
Db 63 -----ProGluAsnLeuAlaAsnSerLeuAspSerTrp 74  
Qy 1363 CTGGGAACCTCTATTGAGGAATAGAGAAATACCTCGGTGGGATCCATTTCCTTCC 1422  
Db 75 ArgGlyValSerMetGluAlaIleHisArg---AsnArgGlnProPheGluLeuGluAsn 93  
Qy 1423 TTAAGACCA---TCTGCAAGTCCACAGTCACTGTTCGGGTAGACCTTCTGAGCAGGA 1479  
Db ---TCTGCAAGTCCACAGTCACTGTTCGGGTAGACCTTCTGAGCAGGA 1479

Db 940 CysArgAsnCysSerThrProGlyProAspIleLeuLeuTyProPheIleTyHisAla 959  
QY 2881 GTTGAGTCAAGTCAGACAGACCACTGACATGCCAGGACAGAGCAGGAC 2931  
Db 960 ValGluSerCysAlaGluThrAlaAspHisSerGlyGlnArgThrGlyThr 976

## RESULT 3

US-09-973-451-2  
; Sequence 2, Application US/09973451  
; Patent No. US20020132328A1  
; GENERAL INFORMATION:  
; APPLICANT: JACOBSON, Myron K.  
; APPLICANT: JACOBSON, Elaine L.  
; APPLICANT: AM, Jean-Christophe  
; APPLICANT: LIN, Winston  
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOSYLASE  
; TITLE OF INVENTION: (PARG) ENZYMES,  
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIVE  
; TITLE OF INVENTION: THEREWITH  
; FILE REFERENCE: NIAD 201  
; CURRENT APPLICATION NUMBER: US/09/973,451  
; CURRENT FILING DATE: 2001-10-09  
; PRIOR APPLICATION NUMBER: US/09/302,812  
; PRIOR FILING DATE: 1999-04-30  
; PRIOR APPLICATION NUMBER: 60/083,768  
; PRIOR FILING DATE: 1998-05-01  
; NUMBER OF SEQ ID NOS: 38  
; SEQ ID NO 2  
; LENGTH: 977  
; TYPE: PRT  
; ORGANISM: Bos taurus  
; FEATURE:  
US-09-973-451-2

## Alignment Scores:

Pred. No.: 0 Length: 977  
Score: 4332.50 Matches: 820  
Percent Similarity: 89.86% Conservative: 57  
Best Local Similarity: 84.02% Mismatches: 90  
Query Match: 63.97% Indels: 9  
DB: 9 Gaps: 5

US-09-302-812-5 (1-3814) x US-09-973-451-2 (1-977)

QY 28 ATGAGTGGGGCCCGCTGGGCGCCCTGCAGGAA--CGCGCTGGGCGCGCTGGA 84  
Db 1 MetSerAlaGlyProGlyCysGluProCysThrLysArgProArgTrpAspAlaAla 20  
QY 85 ACTTCTGGCGGCGACTGCTCGGACTCCCGGAGCTTCCCTGGCAGGACAGGCGGTCTTC 144  
Db 21 ThrSerProAlaAlaSerAspAlaArgSerPheProGlyArgGlnArgValLeu 40  
QY 145 GACCCCAAGGAGCTCCGCTCAGTTGAGGTCCTCTCCCTCCCTCGCAGCCTCGGCTCG 204  
Db 41 AspSerLysAspAlaProValGlnPheArgValProProSerSerSerGlyCysAlaLeu 60  
QY 205 GGGCGGGGGGACCCGACAGGCGACCGCCACTCGTTGTTGTTTCAAAAGAGCTATT 264  
Db 61 GlyArgAlaGlyGlnHisArgGlySerAlaThrSerLeuValPheLysGlnLysThrIle 80  
QY 265 ACTACTGGATGATACATAAGGACCCCAAGACAGCTGAATCAGAA-----AGTAA 315  
Db 81 ThrSerTrpMetAspThrLysGlyIleLysThrValGluSerGluSerLeuHisSerLys 100  
QY 316 GAAACAACAATACAGAATTGACTCCATGATGAGTTCTGTGCGAGAGATCACTTTTAC 375  
Db 101 GluAsnAsnThrArgGluGluSerMetMetSerValGlnLysAspAsnPheTyr 120  
QY 376 CCACATAAGTGGAAAAATTGAAATGTTCTCTAGCTAAATCTTGATAAATCACCACA 435  
Db 121 GlnHisAsnMetGluLysLeuGluAsnValSerGlnLeuGlyPheAspLysSerProVal 140  
QY 436 GAAAGAGTTTACAGTATTGAAACCAACAGCAGACTGCGAGTGTGTGCAAGTGGCAGAT 495

Db 141 GluLysGlyThrGlnTyLeuLysGlnHisGlnThrAlaAlaMetCysLysTrpGlnAsn 160  
QY 496 GAAGGAAGCATGCGAGAACAGCTTTTGGCAAGTGAAGCTCCCGCGGGAGCTCCGCTACCA 555  
Db 161 GluGlyProHisSerGluArgLeuLeuGluSerGluProProAlaValThrLeuValPro 180  
QY 556 AAGCAGCTTAGTAATCTTAACATTTGGTCACTCACCACACACCTGATGACACACAGTACACA 615  
Db 181 GluGlnPheSerAsnAlaAsnValAspGlnSerSerProLysAspHisSerAspThr 200  
QY 616 GATCATGAAGAAGACAGACACCAATCAGCAGATTTTACACCTATAAACTTCCAAATACA 675  
Db 201 AsnSerGluGluSerArgAspAsnGlnPheLeuThrHisValLysLeuAlaAsnAla 220  
QY 676 AAGCAACAGTAGTAGAGATGGGCGAG-----GCCAGAACCACTGTAAGTGCAGTGA 726  
Db 221 LysGlnThrMetGluAspGluGlnArgGluAlaArgSerHisGlnLysCysGlyLys 240  
QY 727 TCTCGCCAGTCTGTGAAGAGACTGTACAGCTCTCAACAGGAGGAGTGTGATGTGTACCA 786  
Db 241 AlaCysHisProAlaGluAlaCysAlaGlyCysGlnGlnGluThrAspValValSer 260  
QY 787 GAGAGTCTCTTTCTCAGATGTTGTCGCCAGGACATTTGGAACCTGGACCAAAAATGACAC 846  
Db 261 GluSerProLeuSerAspThrGlySerGluAspValGlyThrGlyLeuLysAsnAlaAsn 280  
QY 847 AAATTGACTGGACAAGAACAGCCTAGTGTATTCCTCCATTTTCAGAAAGAAAGTGAG 906  
Db 281 ArgLeuAsnArgGlnGluSerSerLeuGlyAsnSerProProPheGluLysGluSerGlu 300  
QY 907 CTGAGTACCAATGATGTACAGCACTCGAGAAACAGTTGTCAAGATTCAAGATTCAGAACGAG 966  
Db 301 ProGluSerProMetAspValAspAsnSerLysAsnSerCysGlnAspSerGluAlaAsp 320  
QY 967 GAAGAAACAAAGTCCAGTCTTTGATCAGCAAGATGATCGTTCTCTCC---CAAAACAGCAAA 1023  
Db 321 GluGluThrSerProGlyPheAspGluGlnGluAspSerSerSerAlaGlnThrAlaAsn 340  
QY 1024 AAATTTCAAGTTGCCAAGCAAGAGCTGATGCGGATCTTTAGGAAACGGTATTGACT 1083  
Db 341 LysProSerArgPheGlnProArgGluAlaAspThrGluLeuArgLysArgSerAla 360  
QY 1084 AAGGAAAGTGAAGTTAGATTGCAATTCCAATTTGAA--GGAGAAATAATCTCGGAC 1140  
Db 361 LysGlyGlyGluLeuArgLeuHisPheGlnPheGluGlyGlyGluSerArgAlaGlyMet 380  
QY 1141 AGTGACTTAAATGCCAAGCCATCTCGAACTCTTCTAGCCTTAATGTAGAGTGTAGAAT 1200  
Db 381 AsnAspValAsnAlaLysArgProGlySerThrSerSerLeuAsnValGluCysArgAsn 400  
QY 1201 TCCAGCGCATGGAAAAAGGAGTCTTAAATATACAGATCAATTTTCATGAAATTTCCAAG 1260  
Db 401 SerLysGlnHisGlyArgLysAspSerLysIleThrAspHisPheMetArgValProLys 420  
QY 1261 TCAGAGGACAGAGAAAGAACATGTGAAGTTCAGACATCAAGACAGAGAGAGATT 1320  
Db 421 AlaGluAspLysArgLysGluGlnCysGluMetLysHisGlnArgThrGluArgLysIle 440  
QY 1321 CCAAATATATCCCACTAACCTCCCTCCAGAGAGAGAGTGGCTGGGAACCTCTATTGAG 1380  
Db 441 ProLysTyrlleProProHisLeuSerProAspLysLysTrpLeuGlyThrProIleGlu 460  
QY 1381 GAAATGAGAAAAATCCCTCGGTGGGATCCATTTGCTTCCCTTCCCTTCCCTTCCCTTCCCT 1440  
Db 461 GluMetArgMetProArgCysGlyIleArgLeuProProLeuArgProSerAlaAsn 480  
QY 1441 CACACAGTGTGTTTCGGGTAGACTTCTGAGCAGCAGGAGAGGTTCCCAAACTTTTCCA 1500  
Db 481 HisThrValThrIleArgValAspLeuLeuArgIleGlyGluValProLysArgProPhePro 500  
QY 1501 ACACATTACAAAGATTCTGTGGATAACAAACATGTGAAATGCGCTTGTTCGGAACAAAAC 1560



QY 676 RAGCCACAGTAGGAGATGGCAGGCCAG-----AGCAACTGTGAAGTGCAGTGA 726  
 Db 220 LyeGlnThrThrGluAspGluHisAlaArgGluAlaLysSerHisGlnLysCysSerLys 239  
 QY 727 TCTCGCCAGTCTGTGAAGAACTGTACAGGCTGTCAACAGAGAGAGTGTGATGTCTACCA 786  
 Db 240 SerCysHisProGlyGluAspCysAlaSerCysGlnAspGluLeuLeuAspValValPro 259  
 QY 787 GAGAGTCTTTGTGACATGTTGTGCGGAGGACATTTGGAACTGGACCAAAAAATGACAAC 846  
 Db 260 LysSerProLeuSerAspValGlySerGluAspValGlyThrGlySerLysAsnAspAsn 279  
 QY 847 AAATGACTGGACAAAGACGCTAGGTGATTCGCTCCATTTTCAGAAAGAAAGTGTAG 906  
 Db 280 LysLeuLeuLeuArgGlnGluSerCysLeuGlyAsnSerProProPheGlnLysGluSerGlu 299  
 QY 907 CTGTAGTACCAATGATGTAGACAACTCGAGAAACAGTTGTCAAGATTCAAGAAGCAGAT 966  
 Db 300 ProGluSerProMetAspValAspAsnSerLysAsnSerCysGlnAspSerGluAlaAsp 319  
 QY 967 GAAGAAACAGTCCAGTCTTTGATGAGCAAGATGAT---CGTTCCTCCCAACAGCAAT 1023  
 Db 320 GluGluThrSerProGlyPheAspGluGlnGluAspGlySerSerGlnThrAlaAsn 339  
 QY 1024 AAATTTCAAGTTGCCAACGACAGAGCTGATGGCGATCTTAGGAAACGCTATTTGACT 1083  
 Db 340 LysProSerArgPheGlnAlaArgAspAlaAspIleGluPheArgLysArgTyrSerThr 359  
 QY 1084 AAGGGAAGTGAAGTTAGATTGCTATTTCCAAATTTGAA---GGAGAAATATGCTGGACC 1140  
 Db 360 LysGlyGlyGluValArgLeuHisPheGlnPheGluGlyGlyGluSerArgThrGlyMet 379  
 QY 1141 AGTGACTTAATGCCAGCCACTGTGAAACTCTTCTAGCTTAATGTAGATGTAGAGT 1200  
 Db 380 AsnAspLeuAsnAlaLysLeuProGlyAsnIleSerSerLeuAsnValGluCysArgAsn 399  
 QY 1201 TCCAGACGATGGAAAAAGGATCTCTAAATTTACAGATCATTTCTATGAGATTTCCAG 1260  
 Db 400 SerLysGlnHisGlyLysAspSerLysIleThrAspHisLeuMetArgLeuProLys 419  
 QY 1261 TCAGAGGACAGAGAAAGAACAAATGTGAAGTCAGACATCAAGAACAGAAAGGAGATT 1320  
 Db 420 AlaGluAspArgArgLysGluGlnTrpGluThrLysHisGlnArgThrGluArgLysIle 439  
 QY 1321 CCAAAATACATCCCACTTAACCTCCCTCCAGAGAGAGTGTGCTGGAACTCTATTGAG 1380  
 Db 440 ProLysTyrValProProHisLeuSerProAspLysLysTrpLeuGlyThrProIleGlu 459  
 QY 1381 GAAATGAGAAAAATGCCTCGGTGTGGATCCATTTCCTTAAGACCATCTCAAGT 1440  
 Db 460 GluMetArgArgMetProArgCysGlyIleArgLeuProLeuLeuArgProSerAlaAsn 479  
 QY 1441 CACACAGTGACTGTTGGGTAGACCTTCTGAGAGAGAGAGTTCCGAAACCTTTTCCA 1500  
 Db 480 HisThrValThrIleArgValAspLeuLeuArgAlaGlyGluValProLysProPhePro 499  
 QY 1501 ACACATTACAAGATTGTGGATACACAACTGTGAAATGCTCTGTTTCGAAACAAAC 1560  
 Db 500 ThrHisTyrLysAspLeuTrpAspAsnLysHisValLysMetProCysSerGluGlnAsn 519  
 QY 1561 TTGTACCTGTGGAAAGATGAGAAATGGTGAAGCACTCGAGGAGTAGTGGAGCTCAT 1620  
 Db 520 LeuTyrProValGluAspGluAsnGlyGluArgThrAlaGlySerArgTrpGluLeuIle 539  
 QY 1621 CAGACTGCATCTTCAACAAATTCACAGCCCGAGAACTTGAAGGATCGCATTCGAAA 1680  
 Db 540 GlnThrAlaLeuLeuAsnLysPheThrArgProGlnAsnLeuLysAspAlaIleLeuLys 559  
 QY 1681 TACATGTGGCATATCTTAAGAAATGGGACTTTACAGCTTTGGTTGATTTCTGGATAG 1740  
 Db 560 TyrAsnValAlaTyrSerLysLysTrpAspPheThrAlaLeuIleAspPheTrpAspLys 579  
 QY 1741 GTACTTGAAGACGAGAGGCCCAACATTTATATCATCAGTCCATTTTACCTGACATGGTGA 1800

Db 580 ValLeuGluGluAlaGluAlaGlnHisLeuTyrGlnSerIleLeuProAspMetValLys 599  
 QY 1801 ATTGCACCTGTCTGCTCCAAATATTTCCACCCAGCCCAATACCACTCTCTGAAAACAGAGATG 1860  
 Db 600 IleAlaLeuCysLeuProAsnIleCysThrGlnProIleProLeuLeuLysGlnLysMet 619  
 QY 1861 AATCATTTCTGTACCATGTACAGAAACAGATCGCGAGTCTTTTAGCTTAATGCTTTCTTC 1920  
 Db 620 AsnHisSerIleThrMetSerGlnGluGlnIleAlaSerLeuLeuAlaAsnAlaPhePhe 639  
 QY 1921 TGCACATTTCCCGCAGCGAATGCCAAGATGAATCGAGTATTTCTAGTTTACCACAGACATT 1980  
 Db 640 CysThrPheProArgArgAsnAlaLysMetLysSerGluTyrSerSerTyrProAspIle 659  
 QY 1981 AACTTCAATCGGTGTCTTTGAAGCAGCTTCATCAAGGAAACCAAGAAACCTGAAAACACTC 2040  
 Db 660 AsnPheAsnArgLeuPheGluGlyArgSerSerArgLysProGluLysLeuLysThrLeu 679  
 QY 2041 TTCCTGCTACTTTCGAAGAGTTCACAGAGAAACCTACAGGATTCGTGACATTTACAGA 2100  
 Db 680 PheCysTyrPheArgArgValThrGluLysLysProThrGlyLeuValThrPheThrArg 699  
 QY 2101 CAGAGTCTTGAAGATTTCCAGAAATGGGAAAGGTGTGAAAGCCTCTGACACGCTTACAC 2160  
 Db 700 GlnSerLeuGluAspPheProGluTrpGluArgCysGluLysProLeuThrArgLeuHis 719  
 QY 2161 GTCACTTACAGGGTACCATAGAACCAAGCCCGAGGATCTCTACAGGTGATTTTGA 2220  
 Db 720 ValThrTyrGluGlyThrIleGluGluAsnGlyGlnGlyMetLeuGlnValAspPheAla 739  
 QY 2221 AATCGTTTGTTCGAGGTGGTGTGACTGTCGGGAGCTTGTACAGAGAAATACAGATT 2280  
 Db 740 AsnArgPheValGlyGlyValThrSerAlaGlyLeuValGlnGluGluIleArgPhe 759  
 QY 2281 TTAATCAATCTCGAATTCATTGTTTCAGCGTGTTCCTGAGGTGCTGAGTGCACATCAG 2340  
 Db 760 LeuIleAsnProGluLeuIleSerArgLeuPheThrGluValLeuAspHisAsnGlu 779  
 QY 2341 TGTCTTATTATCACAGGTACTGAAACAGTACAGTCAATACACAGGCTATCTGTAACCTTAT 2400  
 Db 780 CysLeuIleIleThrGlyThrGluGlnTyrSerGluTyrThrGlyTyrAlaGluThrTyr 799  
 QY 2401 GGTGGGCGCCAGCCATGAAGATGGAGTGAAGAAAGCAGATTGGCAGCGCGCTGCAG 2460  
 Db 800 ArgTrpSerArgSerHisGluAspGlySerGluArgAspAspCysGluArgArgCysThr 819  
 QY 2461 GAGATCGTTGCAATTCAGCCTTACCTTCAGACGCTACCTCGATCAGTTTGTGCTGCTGAG 2520  
 Db 820 GluIleValAlaIleAspAlaLeuHisPheArgArgTyrLeuAspGlnPheValProGlu 839  
 QY 2521 AAAGTGACGTGAGCTTAAACAAGCTTACTCGCGATTCTCTCCGTCCTGAGTTCCTTCT 2580  
 Db 840 LysMetArgArgGluLeuAsnLysAlaTyrCysGlyPheLeuArgProGlyValSerSer 859  
 QY 2581 GAAATCTTTCTGAGTGGCCACCGGAAACCTGGGCTGTGTGCTGCTTGGGGGTGACGCT 2640  
 Db 860 GluAsnLeuSerAlaValAlaThrGlyAsnTrpGlyCysGlyAlaPheGlyGlyAspAla 879  
 QY 2641 AGATTAAAAGCTTAAATACAGATCTGCGAGCTCTCGCGCTGAACCTGAGCTGCTTTAT 2700  
 Db 880 ArgLeuLysAlaLeuIleGlnIleLeuAlaAlaAlaAlaGluArgAspValValTyr 899  
 QY 2701 TTCACCTTTGGGACTCAGAGTTGATGAGAGACATTTTACAGCATGCACACTTCTTACC 2760  
 Db 900 PheThrPheGlyAspSerGluLeuMetArgAspIleTyrSerMetHisIlePheLeuThr 919  
 QY 2761 CAGAGGAGCTGGATGTTGGAAAGTGTACAGTATTGTCTAGACTACTACATGAAGAA 2820  
 Db 920 GluArgLysLeuThrValGlyAspValTyrLysLeuLeuLeuArgTyrTyrAsnGluGlu 939  
 QY 2821 TGCAAGAACTGTTCCACCTCGGACGACAGATCAAGCTTTTATCCATTCATATACCATGCT 2880

```

QY 2008 TCATCAAGGAACACCAAGAACTGAAACACTCTTCTGTACTCTTCAAGAGTCACAG 2067
Db |||||
QY 661 SerSerArgLysProGluLysLeuLysThrLeuPheCysTyrPheArgValThrGlu 680
Db |||||
QY 2068 AAAAAACCTACAGGATTGGTACATTTCACAGACAGAGCTTCAAGATTTCAGAAATGG 2127
Db |||||
QY 681 LysLysProThrGlyLeuValThrPheThrArgGlnSerLeuGluAspPheProGluTyr 700
Db |||||
QY 2128 GAAAGGTGAAAGCTCTGACACGCTTACAGCTACAGTCTTACAGAGGTACATAGAGGC 2187
Db |||||
QY 701 GluArgCysGluLysProLeuThrArgLeuHisValThrTyrGluGlyThrLeuGluGly 720
Db |||||
QY 2188 AACGGCCGAGGATGCTACAGGTGGATTGCAAACTGTTTCTGAGGTGGTGGACT 2247
Db |||||
QY 721 AsnGlyArgGlyMetLeuGlnValAspPheAlaAsnArgPheValGlyGlyValThr 740
Db |||||
QY 2248 GGTGCGGAGCTTGTACAAGAGAAATCAGATTTCATCAATCTGAATTCATGTTTCA 2307
Db |||||
QY 741 GlyAlaGlyLeuValGlnGluGluLeuArgPheLeuLeuAsnProGluLeuValSer 760
Db |||||
QY 2308 CGGCTGTTCACTGAGGTGCTGGATCAATGAGTGTCTTATATCAAGTACTGACAG 2367
Db |||||
QY 761 ArgLeuPheThrGluValLeuAspHisAsnGluCysLeuLeuLeuThrGlyThrGluGln 780
Db |||||
QY 2368 TACAGTGAATACACAGGCTATGCTGAAACTTATCGTTGGCCCGAAGCCATGAAGTGG 2427
Db |||||
QY 781 TyrSerGluTyrThrGlyTyrAlaGluThrTyrArgTyrAlaArgSerHisGluAspGly 800
Db |||||
QY 2428 AGTGAAGAGGAGTGGCAGCGGCGCTGCACGAGATCGTTGCCATTCAGCCTTAC 2487
Db |||||
QY 801 SerGluLysAspAspTrpGlnArgArgCysThrGluLeuValAlaLeuAspAlaLeuHis 820
Db |||||
QY 2488 TTCAGAGCTACTCTGATCAGTTGTCCTGCAAGAGTGCAGAGTGCAGCTTAACAGGCT 2547
Db |||||
QY 821 PheArgArgTyrLeuAspGlnPheValProGluLysValArgArgGluLeuAsnLysAla 840
Db |||||
QY 2548 TACTGGGATTCCTCGTCTCGAGTTCCTCTGAAATCTCTTCTCAGTGGCCACGGGA 2607
Db |||||
QY 841 TyrCysGlyPheLeuArgProGlyValProSerGluAsnLeuSerAlaValAlaThrGly 860
Db |||||
QY 2608 AACTGGGCTGTGGTCCCTTTGGGGTGACGCTAGATTAAAGCCCTTAATACAGATCCTG 2667
Db |||||
QY 861 AsnTrpGlyCysGlyAlaPheGlyGlyAspAlaArgLeuLysAlaLeuLeuGlnLeu 880
Db |||||
QY 2668 CGAGCTGTCGCGCTGAAGCTGACGTGGTGTATTTTACCTTTGGGACTCAGAGTTGATG 2727
Db |||||
QY 881 AlaAlaAlaAlaAlaGluArgAspValValTyrPheThrPheGlyAspSerGluLeuMet 900
Db |||||
QY 2728 AGACATTTTACAGCATGCACACTTTCCTTACCAGAGGAAGCTGGATGTTGAAAAGTG 2787
Db |||||
QY 901 ArgAspIleTyrSerMetHisThrPheLeuThrGluArgLysLeuAspValGlyLysVal 920
Db |||||
QY 2788 TACAAGTTATTCCTTAGATCTACATGAAGTGAAGTGAAGAACTGTTCCACCCCTGGAACA 2847
Db |||||
QY 921 TyrLysLeuLeuLeuArgTyrTyrAsnGluGluCysArgAsnCysSerThrProGlyPro 940
Db |||||
QY 2848 GACATCAAGCTTTATTCATATACATGCTGTGTGAGTCAAGTCCAGAGCCACTGAC 2907
Db |||||
QY 941 AspIleLysLeuTyrProPheIleTyrHisAlaValGluSerSerAlaGluThrThrAsp 960
Db |||||
QY 2908 ATGCCAGGACAGAAGCGAGCCACC 2931
Db |||||
QY 961 MetProGlyGlnLysAlaGlyThr 968

```

RESULT 2

```

US-09-973-451-4
; Sequence 4, Application US/09973451
; Patent No. US20020132328A1
; GENERAL INFORMATION:
; APPLICANT: JACOBSON, Myron K.
; APPLICANT: JACOBSON, Elaine L.
; APPLICANT: AM, Jean-Christophe
; APPLICANT: LIN, Winston

```

```

; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE
; TITLE OF INVENTION: (PARG) ENZYMES,
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV
; FILE REFERENCE: THEREWITH
; CURRENT APPLICATION NUMBER: US/09/973,451
; PRIOR FILING DATE: 2001-10-09
; PRIOR APPLICATION NUMBER: US/09/302,812
; PRIOR FILING DATE: 1999-04-30
; PRIOR APPLICATION NUMBER: 60/083,768
; NUMBER OF SEQ ID NOS: 38
; SEQ ID NO 4
; LENGTH: 976
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
US-09-973-451-4

```

Alignment Scores:

```

Pred. No.: 0 Length: 976
Score: 4395.00 Matches: 837
Percent Similarity: 90.99% Conservative: 52
Best Local Similarity: 85.67% Mismatches: 78
Query Match: 64.89% Indels: 10
DB: 9 Gaps: 6

```

US-09-302-812-5 (1-3814) x US-09-973-451-4 (1-976)

```

QY 28 ATGAGTGGCGGCGCGCTGGAGCCCTGCACGAAAGCG---CGCTGGGCGCGCTGGA 84
Db |||||
QY 1 MetAsnAlaGlyProGlyCysGluProCysThrLysAlaThrArgTrpGlyAlaAla--- 19
Db |||||
QY 85 ACTTCGCGCCGACTCCCTCGGACTCCCGAGCTCCCTGGCAGGAGCGGTGTTCTC 144
Db |||||
QY 20 ThrThrSerProAlaAlaSerAspAlaArgSerPheProSerArgGlnArgValLeu 39
Db |||||
QY 145 GACCCCAAGAGCTCCCGTCCAGTTCAGGTCCTCCCTCCCTCCGACGCTGCTCG 204
Db |||||
QY 40 AspProLysAspAlaHisValGlnPheArgValProProSerSerProAlaCysValPro 59
Db |||||
QY 205 GCGCGGCGGCGCGCCGACAGAGGCAACGCCACTCGTGTGTTTTCACAAAGAGCTATT 264
Db |||||
QY 60 GlyGlnAlaGlyGlnHisArgGlySerAlaThrSerLeuValPheLysGlnLysThrile 79
Db |||||
QY 265 ACTACTTGATGGATCTAAAGGACCCAGACAGCTGATCAGAA-----AGTAAA 315
Db |||||
QY 80 ThrSerTrpMetAspThrLysGlyIleLysThrAlaGluSerGluSerLeuAspSerLys 99
Db |||||
QY 316 GAAACACAACTACAGAAATTCACCTCCATGAGTCTCTGTCAGAAAGATACCTTTAC 375
Db |||||
QY 100 GluAsnAsnAsnThrArgIleGluSerMetSerSerValGlnLysAspAsnPheTyr 119
Db |||||
QY 376 CCACATAAGTGGAAAAATTTGAAAATGTTCTCAGCTAAATCTTGATAAAATCACCCACA 435
Db |||||
QY 120 GlnHisAsnValGluLysLeuValAsnValSerGlnLeuSerLeuAspLysSerLeuThr 139
Db |||||
QY 436 GAAAGAGATTTCAGTATTGAAACCAACAGACAGCTGCGAGTGTGTGAGTGGCAGAAAT 495
Db |||||
QY 140 GluLysSerThrGlnTyrLeuAsnGlnHisGlnThrAlaAlaMetCysLysTrpGlnAsn 159
Db |||||
QY 496 GAAAGGAGCATGCAGAACGCTTTTGGCAGCTGAGCTCCCGCGGGGAGCTCCGTACCA 555
Db |||||
QY 160 GluGlyLysHisThrGluGlnLeuLeuGluSerGluProGlnThrValThrLeuValPro 179
Db |||||
QY 556 AAGCAGCTTAGTAATGCTAAACATTGCTCAGTCACCCACACTGATGACACAGTGCACACA 615
Db |||||
QY 180 GlnGlnPheSerAsnAlaAsnIleAspArgSerProGlnAsnAspAspHisSerAspThr 199
Db |||||
QY 616 GATCATGAAGAGACAGACAGCAATCAGCAGTTCCTTACACCTATATAAACTTGCAATACA 675
Db |||||
QY 200 AspSerGluGluAsnArgAspAsnGlnGlnPheLeuThrThrValLysLeuAlaAsnAla 219
Db |||||

```

; ORGANISM: Mus musculus

; FEATURE:

US-09-973-451-6

## Alignment Scores:

Pred. No.: 0 Length: 968  
 Score: 5164.00 Matches: 968  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 76.24% Indels: 0  
 DB: 9 Gaps: 0

US-09-302-812-5 (1-3814) x US-09-973-451-6 (1-968)

QY 28 ATGAGTGGGGCCCGCTGGAGCCCTCCAGAAAGCCGCTGGGGCCCGCTGGAAT 87  
 Db 1 MetSerAlaGlyProGlyTrpGluProCysThrLysAlaArgTrpGlyAlaAlaGlyThr 20  
 QY 88 TCTGGCGCCACTGCTCGGACTCCCGAGCTCCCTGGCAGGCAGAGGGGTCTTCGAC 147  
 Db 21 SerAlaProThrAlaSerAspSerArgSerPheProGlyArgGlnArgValLeuAsp 40  
 QY 148 CCCAAGGACGCTCCCGTCCAGTTCAGGTCCTCCGCTCCCTCCAGCCCTGCTCGGG 207  
 Db 41 ProLysAspAlaProValGlnPheArgValProProSerSerProAlaCysValSerGly 60  
 QY 208 CGGGCGGACCCACAGAGGCAACGCCACTCGTTGTTTTCAAACAAAGACTATTACT 267  
 Db 61 ArgAlaGlyProHisArgGlyAsnAlaThrSerPheValPheLysGlnLysThrIleThr 80  
 QY 268 ACTTGGATGGTACTAAAGGACCCAGACAGCTGAATCAGAAAGTAAAGAAACCAACAT 327  
 Db 81 ThrTrpMetAspThrLysGlyProLysThrAlaGluSerGluSerLysGluAsnAsn 100  
 QY 328 ACAAGATTTGACTCCATGATGAGTCTGTGCAAGAAAGTAACTTTTACCACATAAGGTG 387  
 Db 101 ThrArgIleAspSerMetMetSerValGlnLysAspAsnPheTyProHisLysVal 120  
 QY 388 GAAAAATTTGGAAAAATGTTCCCTCAGCTAACTCTGATTAATCAACCCACAGAAAGAGTTCA 447  
 Db 121 GluLysLeuGluAsnValProGlnLeuAsnLeuAspLysSerProThrGluLysSerSer 140  
 QY 448 CAGTATTTGAACCAACAGCAGACTCCGAGTGTGTCAGAGTGCAGAAATGAAGGAGCAT 507  
 Db 141 GlnTyLeuAsnGlnGlnThrAlaSerValCysLysTrpGlnAsnGluGlyLysHis 160  
 QY 508 GCAGAACGCTTTTGGCAAGTGAGCTCCCGCGGAGCTCCGCTACCAAGACAGCTTAGT 567  
 Db 161 AlaGluGlnLeuLeuAlaSerGluProProAlaGlyThrProLeuProLysGlnLeuSer 180  
 QY 568 AATGCTAACTTTGTCAGTCAACCCACACTGATGACACAGTGCAGATCATGAGAA 627  
 Db 181 AsnAlaAsnIleGlyGlnSerProHisThrAspAspHisSerAspThrAspHisGluGlu 200  
 QY 628 GACAGAGCAATCAGCAGTCTTTTACACTATAAACTTGCAAAATGCAAAAGCCAACTGTA 687  
 Db 201 AspArgAspAsnGlnGlnPheLeuThrProIleLysLeuAlaAsnThrLysProThrVal 220  
 QY 688 GGAGATGGGAGCCAGAGCAACTGTAAGTGCAGTGAATCTCGCAGTCTGTGAAAGAC 747  
 Db 221 GlyAspGlyGlnAlaArgSerAsnCysLysCysSerGlySerArgGlnSerValLysAsp 240  
 QY 748 TGTACAGGCTGTCAACAGGAGGAGTGGATGCTTACCAGAGATGCTTCTGTCAGATGTT 807  
 Db 241 CysThrGlyCysGlnGlnGluValAspValLeuProGluSerProLeuSerAspVal 260  
 QY 808 GGTCCCGAGGACATTTGAACTGGACCAAAAAATGACAAATTTGACTGGCAAGAAAGC 867  
 Db 261 GlyAlaGluAspIleGlyThrGlyProLysAsnAspAsnLysLeuThrGlyGlnGluSer 280  
 QY 868 AGCTAGTGTGATTCGCTCCATTGAGAAAGAGTGAAGCTGAGTCAACCAATGGATGTA 927  
 Db 281 SerLeuGlyAspSerProPheGluLysGluSerGluProGluSerProMetAspVal 300

QY 928 GACAACTCGAGAAAACAGTTGTCAAGATTCAAGAGCAGATGAAGAAACAAGTCCAGTCTTT 987  
 Db 301 AspAsnSerArgAsnSerCysGlnAspSerGluAlaAspGluGluThrSerProValPhe 320  
 QY 988 GATGAGCAAGATGATCGTTCTCCCAAAACAGCAAAATAAATTTCAAGTTGCCCAACAGA 1047  
 Db 321 AspGluGlnAspAspArgSerSerGlnThrAlaAsnLysLeuSerSerCysGlnAlaArg 340  
 QY 1048 GAAAGCTGATGGCGATCTTTAGGAAAACGGTATTTGACTAAGGGAAGTGAAGTTAGATTGCAT 1107  
 Db 341 GluAlaAspGlyAspLeuArgLysArgTyLeuThrLysGlySerGluValArgLeuHis 360  
 QY 1108 TTCCAATTTGAAGGAGAAAATAATCTCGGACCAATGACTTAATGCAAGCCATCTGGA 1167  
 Db 361 PheGlnPheGluGlyGluAsnAsnAlaGlyThrSerAspLeuAsnAlaLysProSerGly 380  
 QY 1168 AACTCTTCTAGCCTTAATGTAGAGTGTAGAGTTCACAGCAGCATGGAAGAAAGGATTTCT 1227  
 Db 381 AsnSerSerSerLeuAsnValGluCysArgSerSerLysGlnHisGlyLysArgAspSer 400  
 QY 1228 AAAATTACAGATCAATTTTCATGAGAAATTTCCAAGTCAGAGGACAGAAAGAAACAATGT 1287  
 Db 401 LysIleThrAspHisPheMetArgLysSerLysSerGluAspArgArgLysGluGlnCys 420  
 QY 1288 GAAGTCAGACATCAAGAAACAGAAAGAAAGTTCCAAATACATCCCACTTAACCTCCCT 1347  
 Db 421 GluValArgHisGlnArgThrGluArgLysIleProLysTyIleProProAsnLeuPro 440  
 QY 1348 CCAGAGAAAGAGTGGCTGGAACTCTTATTCAGGAATTCAGAAATTCAGAAATTCGCTCGTGGG 1407  
 Db 441 ProGluLysLysTrpLeuGlyThrProLysGluMetArgLysMetProArgCysGly 460  
 QY 1408 ATCCATTTGCCCTTCTTAAAGACCATCTGCAAGTCACAGTGCAGTGTTCGGGTAGACCTT 1467  
 Db 461 IleHisLeuProSerLeuArgProSerAlaSerHisThrValThrValArgValAspLeu 480  
 QY 1468 CTGAGACGAGAGAGGTTCGAAACCTTTTCCAAACATTAACAAGATTCTGGGATTAAC 1527  
 Db 481 LeuArgAlaGlyGluValProLysProPheProThrHisTyLysAspLeuTrpAspAsn 500  
 QY 1528 AAACATGTGAATTCCTTGTTCGAAACAAACTGTACCTCGTGGAGATGAGAAATGGT 1587  
 Db 501 LysHisValLysMetProCysSerGluGlnAsnLeuTyProValGluAspGluAsnGly 520  
 QY 1588 GAGCAACTGCGAGGATGAGTGGAGTCAATTCAGACTGCATCTCTCAACAAATTCACA 1647  
 Db 521 GluArgThrAlaGlySerArgTrpGluLeuIleGlnThrAlaLeuLeuAsnLysPheThr 540  
 QY 1648 CGACCCAGAACTTGAAGGATGCCATTTCTGAAATACATATGTCATATTTCTAAGAAATGG 1707  
 Db 541 ArgProGlnAsnLeuLysAspAlaIleLeuLysTyAsnValAlaTySerLysLysTrp 560  
 QY 1708 GACTTTACAGCTTTGGTTGATTTCTGGGATAAGTACTTGAAGAGCAGAGGCCCAACAT 1767  
 Db 561 AspPheThrAlaLeuValAspPheTrpAspLysValLeuGluGluAlaGluAlaGlnHis 580  
 QY 1768 TTATATCAGTCCATTTTACCTGACATGGTGAATTCCTCTCTCTCCCAATTTATTCG 1827  
 Db 581 LeuTyGlnSerIleLeuProAspMetValLysIleAlaLeuCysLeuProAsnIleCys 600  
 QY 1828 ACCCAGCCATACCACTTCCTGAAACAGACAGATGATCATCTCTGTCAGTGTTCACAGAA 1887  
 Db 601 ThrGlnProIleProLeuLeuLysGlnLysMetAsnHisSerValThrMetSerGlnGlu 620  
 QY 1888 CAGATCCCAAGTCTTTTGTAGCTTAATGCTTCTTCTGCACTTTCCCGACCGAATGCCAAG 1947  
 Db 621 GlnIleAlaSerLeuLeuAlaAsnAlaPhePheCysThrPheProArgArgAsnAlaLys 640  
 QY 1948 ATGAATTCGGAGTATCTAGTTACCCAGACATTAACCTCAATCGTTGTTTGAAGGACGT 2007  
 Db 641 MetLysSerGluTySerSerTyProAspIleAsnPheAsnArgLeuPheGluGlyArg 660



Db	432	-----AlaArgProAspSerProPheSer-----Phe	440
QY	1987	AATCGGTTGTTGAAGGACGTTCAATCAAGGAACACGAAAACTGAAAAACACTCTTCTGCG	2046
Db	441	CysArgIleLeuSerSerAspLysSerIleCysValGluLysLeuLysPheLeuPheThr	460
QY	2047	TACTTTTCCGAAGAGTCACAGAGAAAAACCTACAGGATGGTGACATTT-----ACAAGA	2100
Db	461	TyrPheAspLysMetSerMetAspProProAspGlyAlaValSerPheArgLeuThrLys	480
QY	2101	CAGAGTCTTGAAGATTTT---CCAGATGG---GAAAGGTGTCAAAGCCTCGACACGC	2154
Db	481	MetAspLysAspThrPheAsnGluGluTrpLysAspLysLysLeuArgSerLeuProGlu	500
QY	2155	TTACACGTCCTACTTACGAGGTACCATAGAAGCAACGGCCGAGGCATGCTACAGGTGGAT	2214
Db	501	ValGluPhePheAspGluMetLeuIleGluAspThrAlaLeu---CysThrGlnValAsp	519
QY	2215	TTTTCRAATCGTTTGTGTGAGGTGGTGTCAGCTGGTGGGACCTTCTACAGAGAAATC	2274
Db	520	PheAlaAsnGluHisLeuGlyGlyGlyValLeuAsnHisGlySerValGlnGluGluIle	539
QY	2275	AGATTTTAAATCAATCCTGAATTGATTGTTTTCACGGTGTTCTCACTGAGGTGCTGCATCAC	2334
Db	540	ArgPheLeuMetCysProGluMetMetValGlyMetLeuLeuCysGluLysMetLysGln	559
QY	2335	AATGAGTGTCTTATTATCAcAGGTACTGGAACAGTACAGTGAATACACAGGCTATGCTGAA	2394
Db	560	LeuGluAlaIleSerIleValGlyAlaTyrValPheSerSerTyrThrGlyTyrGlyHis	579
QY	2395	ACTTATCGTTGGGCC-----CGAAGCCATCAAGATGGAGTGA---AAG	2436
Db	580	ThrLeuLysTrpAlaGluLeuGlnProAsnHisSerArgGlnAsnThrAsnGluPheArg	599
QY	2437	GACGATTGGCAGCGGCGCTCCACGGAGATCGTTGCCATTGACGCACCTTCACTTCAGACGC	2496
Db	600	AspArgPheGlyArgLeuArgValGluThrIleAlaIleAspAlaIleLeuPheLysGly	619
QY	2497	-----TACCTCGATCAGTTTGTGCTCGCTGAGAAAGTGAAGCTGAGCTTAAC	2541
Db	620	SerLysLeuAspCysGlnThrGluGlnLeuAsnLysAlaAlaSerIleLeuArgGluMetLys	639
QY	2542	AAGGCTTACTTGGGATTCCTCCGTCCTGGAGTTCCTTCTGAAAATCTTTCTCGCAGGTGGCC	2601
Db	640	LysAlaSerIleGlyPheMetSerGlnGlyProLysPheThrAsnIle---ProIleVal	658
QY	2602	ACGGAAACATGGGGCTGTGTGCTTTGGGGGTGACGCTAGATTAAGAGCTTAATACAG	2661
Db	659	ThrGlyTrpTrpGlyCysGlyAlaPheAsnGlyAspLysProLeuLysPheIleGln	678
QY	2662	ATCTGSCAGCTGTGGCTGGCTGAACGTGACGTGTTTATTTCACCTTGGGGACTCAGAG	2721
Db	679	ValIleAlaAlaGlyValAlaAspArgProLeuHisPheCysSerPheGlyGluProGlu	698
QY	2722	TTGATGAGACATTTACAGCATGCACATTTCTCTTCCGAGAGGAAGCTGGATGTGA	2781
Db	699	LeuAlaAlaLysCysLysLysIleIleGluArgMetLysGlnLysAspValThrLeuGly	718
QY	2782	AAA	2784
Db	719	Lys	719

Db 719 Lys 719

RESULT 15

US-09-511-507-10

Sequence 10, Application US/09511507

Patent No. 6395543

GENERAL INFORMATION:

APPLICANT: JACOBSON, Myron K.

APPLICANT: JACOBSON, Elaine L.

APPLICANT: AME, Jean-Christophe

APPLICANT: LIN, Winston

TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE (PARG) EN

TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV

TITLE OF INVENTION: THEREWITH

FILE REFERENCE: NIAD 201

CURRENT APPLICATION NUMBER: US/09/511,507

CURRENT FILING DATE: 2000-02-23

PRIOR FILING DATE: 09/302,812

PRIOR FILING DATE: 1999-04-30

NUMBER OF SEQ ID NOS: 38

SEQ ID NO 10

LENGTH: 726

TYPE: PRT

ORGANISM: Caenorhabditis elegans

FEATURE:

US-09-511-507-10

Alignment Scores:

Pred. No.:	1.27e-33	Length:	726
Score:	443.50	Matches:	184
Percent Similarity:	36.62%	Conservative:	124
Best Local Similarity:	21.88%	Mismatches:	278
Query Match:	6.55%	Indels:	255
DB:	4	Gaps:	32

US-09-302-812-5 (1-3814) x US-09-511-507-10 (1-726)

QY 622 GAAGACAGACAGACAATCAGCAGTTCTTACACTATAAACTTGCAATACAAAGCCA 681

Db 14 GlnAspGluLysAspTyrGluAspTyrVal---GlyValGlyPheAlaHisGlnValPro 32

QY 682 ACAGTAGAGATGGCGGCGGAGAGCACTGTAGTGCAGTGGATCGCCAGCTGTGTG 741

Db 33 ThrMet-----LysArgArgLysLeuThrGluHisGlyAsnThrThrGluSerLys 49

QY 742 AAAGCTGTACAGCTGTCAACAGAGGAGGTGTGTGTACAGAGAGTCTTGTCA 801

Db 50 GluAspProGlu-----GluProLysSerArgAspValPheValSerSerGlnSerSer 67

QY 802 GATGTTGTCGCGAGGACATTTGGAACTGGACCAAAA----- 837

Db 68 AspGluSerGlnGluAspSerAlaGluAsnProGluLeuAlaLysGluValSerGluAsn 87

QY 838 -----AATGACAACAAAATTG 852

Db 88 CysGluAsnLeuThrGluThrLeuLysLysSerAsnLeuGluSerLeuAspAsnValThr 107

QY 853 ACTGGACAAGAACAGCCCTAGGTGAT-----TCGCTCCATTTGAGAAAGAAAGT 903

Db 108 GluArgSerGluHisThrLeuAsnAsnHisLysSerThrGluProMetGluGluAspVal 127

QY 904 GAGCTCAGTCAACCAATGGAGTGA----- 927

Db 128 AsnAsnLysSerAsnLeuAspValAlaLeuAsnSerAspGluAspGluValLeu 147

QY 928 ---GACAACTCGAGAACAGTTGTCAAGATTCAGAGACAGATGAGAAACAAAGTCCAGTC 984

Db 148 GluGluAsnAsnLysGluMetArgAspGlyGluGlnValGlnGlnLeuSerGluAspLeu 167

QY 985 TTT---GATGAGCAAGAT----- 999

Db 168 PheAlaAspAspGlnGluLeuLeuGluTyrProGlyLeuMetLysAspThrThrThrGln 187

QY 1000 -----GATCGTTCTCCAAAACAGCAATAATACTTCAAGTTGCCAAGCAAGA 1047

Db 188 LeuAspIleThrAspSerGluValGluThrAlaGlnLysMetGluMetIleGluGluThr 207

QY 1048 GAAGCTGATGCGGATCTTAGGAAACGGTATTTGACTAAGGAAAGTGAAGTTAGATTGCAT 1107

Db 208 GluAlaAspSer----- 211

QY 1108 TTCATTTGAAGGAGAGAAATAATGCTGGGACCAGTCACTTAATGCCAAGCCATCTGGA 1167

Db 212 ---ThrPheValGlyGluAspSerLysAlaThrLysThrValArgThrSerSerSer 230

QY 1168 AACTCTTCTAGCCTTAATGTAGAGTGTAGAAGTTCCAAAGCAGCATGCAAAAAGG----- 1221

Db 231 PheLeuSerThrValSerThr---CysGluAlaProAlaLysGlyArgAlaArgMetTyr 249

QY 1222 GATTCTAAATATACAGATCATTTTCATGAGAAATTTCCAAAGTCAGAG----- 1266

Db 250 GlnLysGluLeuGluLysHisValIleAlaPheThrGluGlyAsnLeuThrLeuGlnPro 269

QY 1267 -----GACAGAGAAAAGACAATGTGAAGTCAGATCAGCATCAAGA 1305

Db 270 AspLeuAsnLysValAspProAspArgAsnTyrArgTyrCysThrIle----- 285

QY 1306 ACAGAAAGGAAGATTCCAAAATACATCCACCTAACTCCCTCCAGAGAAAGAGTGGCTG 1365

Db 286 -----ProAsnPheProAlaSerGln----- 292

QY 1366 GGAACCTCTATTGAGGAAATGAGAAAATGCTCGGTGGATTCATTTGCCTTCCTTA 1425

Db 293 GlyLysLeuArgGluAspAsnArgTyrGlyProLysIleValLeu----- 307

QY 1426 AGACCATCTCGAAGTCACACAGTGTGTTCCGGGTAGACCTTCTGAGAGCAGGAGGTT 1485

Db 307 ----- 307

QY 1486 CCGAAACCTTTTCCACACATTTACAAAGATTTGTGGGATAACAAACATGTGAAAATGCCT 1545

Db 308 -----ProGlnArgTyrArgGluPhe----- 314

QY 1546 TGTTCGGAACAAACTTGTACCTGTGGAGATGAGAAATCGTGAGCGA----- 1593

Db 315 -----AspSerArgGlyArgArgAspSerTyr 324

QY 1594 -----ACTGCAGG----- 1602

Db 325 PheTyrPheLysArgLysLeuAspGlyTyrLeuLysCysTyrLysThrThrGlyTyrPhe 344

QY 1603 -----AGTAGTGGAGCTCATTTCAGACTGCATCTCTCAACAAA 1641

Db 345 MetPheValGlyLeuLeuHisAsnMetTyrGluPhe----- 356

QY 1642 TTCACAGCCCCAGAACCTTGAAGAGTGCATTTCTGAAATACAAATGTGGCA----- 1692

Db 357 -----AspProAspIleThrTyrLysLeuProAlaLeuGlu 368

QY 1693 -----TATTCTAAGAAATCGGACTTTACAGCTTTGTTTGTTCATTTCTGGATAGGTACTT 1746

Db 369 MetTyrTyrLysGluMetSerGluLeuValGlyArgGluGluValLeuGluLysPheAla 388

QY 1747 GAAGAACAGCAGGCCCAACATTTATATACGTCCATTTTACCTGCATGTTGTGAAAATTGCA 1806

Db 389 ArgValAlaArgIleAlaLysThrAlaGluAspIleLeuProGlu-----ArgIleTyr 406

QY 1807 CTCTGCTGCCAAATATTTGCAACCCAGCCCAATACCCTCTCTGAAACAGAGATGAATCAT 1866

Db 407 ArgLeuValGlyAspVal-----Glu 413

QY 1867 TCTGTACAGATGTCACAGGACAGATCGCAGTCTTTTACCTAATGCTTTCTTCTGCACA 1926

Db 414 SerAlaThrLeuSerHisLysGlnCysAlaAlaLeuValAlaArgMetPhePhe----- 431

QY 1927 TTTCCCGCAGGAATGCCAAGATGAAATCGAGTATTTCTAGTTACCCAGACATTAATCTTC 1986



Db 345 MetPheValGlyLeuLeuHisAsnMetTrpGluPhe----- 356  
 Qy 1642 TTCACAGACCCAGAACTTGAAGATCGGATTCTGAATATCAATGTGCA----- 1692  
 Db 357 -----AspProAspIleThrTyrLysLeuProAlaLeuGlu 368  
 Qy 1693 -----TATTCTAAGAAATGGACTTTACAGCTTTGGTTGATTCTCGGATAAGTACTT 1746  
 Db 369 MetTyrTyrLysGluMetSerGluLeuValGlyArgGluValLeuGluLysPheAla 388  
 Qy 1747 GAAGAGCAGAGGCCCAACATTATATCAGTCCATTTTACCTGACATCGTGAATTCGA 1806  
 Db 389 ArgValAlaArgIleAlaLysThrAlaGluAspIleLeuProGlu-----ArgIleTyr 406  
 Qy 1807 CTCTGCTGCCAAATATTGTCACCCAGCCCAATACCACTCTCGAAACAGAGATGAATCAT 1866  
 Db 407 ArgLeuValGlyAspVal-----Glu 413  
 Qy 1867 TCTGTCAGATGTCACAGAAACAGATCGCCAGTCTTTAGCTAATGCTTTCTTCGCACA 1926  
 Db 414 SerAlaThrLeuSerHisLysGlnCysAlaAlaLeuValAlaArgMetPhePhe----- 431  
 Qy 1927 TTTCCCGACGGAATGCAAGATGAATCGAGTATTTCTAGTACCACAGATTAATCTTC 1986  
 Db 432 -----AlaArgProAspSerPheSer-----Phe 440  
 Qy 1987 AATCGGTTGTTGAAGGAGCTTCATCAAGGAAACCCAGAAACAACTGAAACACTCTCTGC 2046  
 Db 441 CysArgIleLeuSerSerAspLysSerIleCysValGluLysLeuLysPheLeuPheThr 460  
 Qy 2047 TACTTTCCGAGAGTCACAGAAACAACTCAGAGATTGGTGACATTT-----ACAAGA 2100  
 Db 461 TyrPheAspLysMetSerMetAspProAspGlyAlaValSerPheArgLeuThrLys 480  
 Qy 2101 CAGAGTCTTGAGATTTT-----CCAGATGG-----GAAAGCTGTGAAGCCCTCTGACAGC 2154  
 Db 481 MetAspLysAspThrPheAsnGluTrpLysAspLysLeuArgSerLeuProGlu 500  
 Qy 2155 TTACAGCTCACTTACGAGGGTACATAGAACCAAGCCGAGGCGATGCTACAGTGGAT 2214  
 Db 501 ValGluPhePheAspGluMetLeuIleGluAspThrAlaLeu---CysThrGlnValAsp 519  
 Qy 2215 TTGCAAAATCGTTTGTGGAGGTGGTGAGTGGTGGGAGCTGTGACAGAGAAATC 2274  
 Db 520 PheAlaAsnGluHisLeuGlyGlyValLeuAsnHisGlySerValGlnGluGluLe 539  
 Qy 2275 AGATTTTAAATCAATCTGATTCATGTTTTCACGGCTGTTCACTGAGTCTGATCAC 2334  
 Db 540 ArgPheLeuMetCysProGluMetMetValGlyMetLeuLeuCysGlyLysMetLysGln 559  
 Qy 2335 AATGAGTGTCTTATATACAGGCTACTGAAACAGTACAGTGAATACAGAGCTATGCTGAA 2394  
 Db 560 LeuGluAlaIleSerIleValGlyAlaTyrValPheSerSerTyrThrGlyTyrGlyHis 579  
 Qy 2395 ACTTATCGTTGGGCC-----CGAAGCCATGAAGATGGAGTGAA---AAG 2436  
 Db 580 ThrLeuLysTrpAlaGluLeuGlnProAsnHisSerArgGlnAsnThrAsnGluPheArg 599  
 Qy 2437 GAGGATGGCAGCGCGCTGACCGAGATCGTTGCCATTGACGCACTTCACTTCAGAGCC 2496  
 Db 600 AspArgPheGlyArgLeuArgValGluThrIleAlaIleAspAlaIleLeuPheLysGly 619  
 Qy 2497 -----TACCTCGATCAGTTTGTGCTCGAAGAGTGACAGCTGAGCTTAAC 2541  
 Db 620 SerLysLeuAspCysGlnThrGluGlnLeuAsnLysAlaAsnIleIleArgGluMetLys 639  
 Qy 2542 AAGCTTACTCGGATTCCTCCGCTCGAGTCTCTTCTGAAATCTTTCTGCAAGTGGCC 2601  
 Db 640 LysAlaSerIleGlyPheMetSerGlnGlyProLysPheThrAsnIle---ProIleVal 658  
 Qy 2602 ACCGGAACCTGGGCTGCTGCTTGGGGGTGACCGCTAGATTAAGCCCTTAATACAG 2661

Db 659 ThrGlyTrpTrpGlyCysGlyAlaPheAsnGlyAspLysProLeuLysPheIleIleGln 678  
 Qy 2662 ATCTCTGCGAGCTGCTGGCTTGAACGTGACCTGCTTTTATTTTCACTTTGGGACTCAGAG 2721  
 Db 679 ValIleAlaAlaGlyValAlaAspArgProLeuHisPheCysSerPheGlyGluProGlu 698  
 Qy 2722 TTGATGAGAGACATTTTACAGCATCAGACATTTCTTCCACGAGAGAGAACTGGATGTTGGA 2781  
 Db 699 LeuAlaAlaLysCysLysLysIleIleGluArgMetLysGlnLysAspValThrLeuGly 718  
 Qy 2782 AAA 2784  
 Db 719 Lys 719  
 RESULT 14  
 ; Sequence 10, Application US/09511477  
 ; Patent No. 6337202  
 ; GENERAL INFORMATION:  
 ; APPLICANT: JACOBSON, Myron K.  
 ; APPLICANT: JACOBSON, Elaine L.  
 ; APPLICANT: AME, Jean-Christophe  
 ; APPLICANT: LIN, Winston  
 ; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE (PARG) EN  
 ; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV  
 ; FILE REFERENCE: NIAD 201  
 ; CURRENT APPLICATION NUMBER: US/09/511,477  
 ; CURRENT FILING DATE: 2000-02-23  
 ; PRIOR APPLICATION NUMBER: 09/302,812  
 ; PRIOR FILING DATE: 1999-04-30  
 ; NUMBER OF SEQ ID NOS: 38  
 ; SEQ ID NO 10  
 ; LENGTH: 726  
 ; TYPE: PRT  
 ; ORGANISM: Caenorhabditis elegans  
 ; FEATURE:  
 US-09-511-477-10  
 Alignment Scores:  
 Pred. No.: 1,27e-33 Length: 726  
 Score: 443.50 Matches: 184  
 Percent Similarity: 36.62% Conservative: 124  
 Best Local Similarity: 21.88% Mismatches: 278  
 Query Match: 6.55% Indels: 255  
 DB: 4 Gaps: 32  
 US-09-302-812-5 (1-3814) x US-09-511-477-10 (1-726)  
 Qy 622 GAAGAAGACAGACAAATCAGCAGTTTCTTACACTATAAACTTGCATAACAAAGCCA 681  
 Db 14 GlnAspGluLysAspTyrGluAspTyrVal---GlyValGlyPheAlaHisGlnValPro 32  
 Qy 682 ACAGTAGGAGATGGGAGGCGGAGCCAGAACACTGTAAGTCAGTCGATCTGCCAGTCTGTG 741  
 Db 33 ThrMet-----LysArgArgLysLeuThrGluHisGlyAsnThrThrGluSerLys 49  
 Qy 742 AAAGACTGTACAGCTCTCAACAGGAGGAGGTGGATGTGTCTACCAGAGAGTCTCTTGTCA 801  
 Db 50 GluAspProGlu-----GluProLysSerArgAspValPheValSerSerGlnSerSer 67  
 Qy 802 GATGTTGGTCCGAGGACATTTGGAAGTGGACCAAAA----- 837  
 Db 68 AspGluSerGlnGluAspSerAlaGluAsnProGluIleAlaLysGluValSerGluAsn 87  
 Qy 838 -----AATGACAAACAATTG 852  
 Db 88 CysGluAsnLeuThrGluThrLeuLysIleSerAsnIleGluSerLeuAspAsnValThr 107  
 Qy 853 ACTGGCAAGAAAGCAGCTTAGTGTAT-----TCGCTCCATTTGAGAAAGAAAGT 903  
 Db 108 GluArgSerGluHisThrLeuAspAsnHisLysSerThrGluProMetGluGluAspVal 127



Db 492 Gly-----ValAlaThrGlyAsnTrpGlyCysGlyAlaPhe 503  
QY 2629 GGSGGTGACCTAGATTAAAGCTTAAATACAGATCCTGGCAGCTGCTGGCTGAACGT 2688  
Db 504 GlyGlyAspSerTyrLeuLysAlaLeuLeuGlnLeuMetValCysAlaGlnLeuGlyArg 523  
QY 2689 GAGCTGTTTATTTCACCTTTGGGACTCAGAGTTGATGAGACACATTTACAGCATGCAC 2748  
Db 524 ProLeuAlaTyrTyrThrPheGlyAsnValGluPheArgAspPheHisGluMetTrp 543  
QY 2749 ACTTCTTACCAGAGGAGCTGGATGTTGGAAGTGTACAGTTATTGCTTAGTAC 2808  
Db 544 LeuLeuPheArgAsnAspGlyThrThrValGlnGlnLeuTrpSer-----IleLeuArgSer 562  
QY 2809 TACAATGAGAATGACGAAACTGTTCACCCCTGGACCA-----GAC 2850  
Db 563 TyrSerArgLeuLeuLysGluLysSerSerLysGluProArgGluAsnLysAlaSerLys 582  
QY 2851 ATCAAGCTTTATCCATTATCATATCCATGCTGTTGAGTCAAGTGACAGACACCTGACATG 2910  
Db 583 LysLysLeuTyrAspPheIle-----LysGluGluLeuLysLysValArgAspVal 599  
QY 2911 CCAGGACAGAGGCA 2925  
Db 600 ProGlyGluGlyAla 604  
RESULT 13  
US-09-302-812-10  
; Sequence 10, Application US/09302812B  
; Patent No. 6333148  
; GENERAL INFORMATION:  
; APPLICANT: JACOBSON, Myron K.  
; APPLICANT: JACOBSON, Elaine L.  
; APPLICANT: AM, Jean-Christophe  
; APPLICANT: LIN, Winston  
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE (PARG) ENZ  
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIVE  
; FILE REFERENCE: NIAD 201  
; CURRENT APPLICATION NUMBER: US/09/302,812B  
; CURRENT FILING DATE: 1999-04-30  
; EARLIER APPLICATION NUMBER: 60/083,768  
; EARLIER FILING DATE: 1998-05-01  
; NUMBER OF SEQ ID NOS: 38  
; SEQ ID NO 10  
; LENGTH: 726  
; TYPE: PRT  
; ORGANISM: Caenorhabditis elegans  
; FEATURE:  
US-09-302-812-10  
Alignment Scores:  
Pred. No.: 1-27e-33 Length: 726  
Score: 443.50 Matches: 184  
Percent Similarity: 36.62% Mismatches: 124  
Best Local Similarity: 21.88% Indels: 255  
Query Match: 6.55% Gaps: 32  
DB: 4  
US-09-302-812-5 (1-3814) x US-09-302-812-10 (1-726)  
QY 622 GAAGAAGACAGACAGCAATCAGCAGTTTCTTACACCTTATAAACTTCCAAATACAAAGCCA 681  
Db 14 GlnAspGluLysAspTyrGluAspTyrVal---GlyValGlyPheAlaHisGlnValPro 32  
QY 682 ACAGTAGGAGATGGGACGCCCAAGCAACTGTAAGTGAGTGAGTCTGCCAGTCTGTG 741  
Db 33 ThrMet-----LysArgArgLysLeuThrGluHisGlyAsnThrThrGluSerLys 49  
QY 742 AAAGACTGTACAGCTGTCAACAGGAGGAGGTGTGCTACACAGAGATCCTTTGTCA 801  
Db 50 GluAspProGlu-----GluProLysSerArgAspValPheValSerSerGlnSerSer 67

QY 802 GATGTTGGTCCGAGGACATTTGGAATCGGACCAAAA----- 837  
Db 68 AspGluSerGlnGluAspSerAlaGluAsnProGluIleAlaLysGluValSerGluAsn 87  
QY 838 -----AATGACAACAATTG 852  
Db 88 CysGluAsnLeuThrGluThrLeuLysIleSerAsnIleGluSerLeuAspAsnValThr 107  
QY 853 ACTGGACAACAGCAGCTTAGTGAT-----TCGCTCCATTGTTGAGAAGAAGT 903  
Db 108 GluArgSerGluHisThrLeuAspAsnHisLysSerThrGluProMetGluGluAspVal 127  
QY 904 GAGCTGAGTCACCAATGGATGA----- 927  
Db 128 AsnAsnLysSerAsnIleAspValAlaIleAsnSerAspGluAspGluLeuValLeu 147  
QY 928 ---GACAACTCGAGAACAGTTGTCAGATTTCAGAGCAGATCAGAAACAAGTCCAGTC 984  
Db 148 GluGluAsnAsnLysGluMetArgGlyGluGlnValGlnGlnLeuSerGlnAspLeu 167  
QY 985 TTT---GATGAGCAAGAT----- 999  
Db 168 PheAlaAspAspGlnGluLeuIleGluTyrProGlyIleMetLysAspThrThrThrGln 187  
QY 1000 -----GATCGTTTCTCCCAACACAGCAAAATAAACTTTCAAGTTCCCAAGCAAGA 1047  
Db 188 LeuAspIleThrAspSerGluValGluThrAlaGlnLysMetGluMetIleGluGluThr 207  
QY 1048 GAGCTGATGGCGATCTTAGGAAACGTTATTCAGTAAAGGAGTGAAGTTAGTATGTCAT 1107  
Db 208 GluAlaAspSer----- 211  
QY 1108 TTCCAATTTGAAGGAGAAATAATGCTGGACCCAGTGACTTAATGCCAAGCCATCTGGA 1167  
Db 212 ---ThrPheValGlyGluAspSerLysAlaThrLysThrValArgThrSerSerSer 230  
QY 1168 AACTCTTCTAGCTTAATGATGAGTGTAGAGTTCCAGCAGCAGCATGCAAAAAGG----- 1221  
Db 231 PheLeuSerThrValSerThr---CysGluAlaProAlaLysGlyArgAlaArgMetTyr 249  
QY 1222 GATTCCTAAATACAGATCATTTTCAGAAATTTCCAAGTCAGAG----- 1266  
Db 250 GlnLysGluLeuGluLysHisValIleAlaPheThrGluGluAsnLeuThrLeuGlnPro 269  
QY 1267 -----GACAGAAGAAAACAATGTGAAGTCAGACATCAAGCA 1305  
Db 270 AspLeuAsnLysValAspProAspArgAsnTyrArgTyrCysThrIle----- 285  
QY 1306 ACAGAAAGGAAGATTCCAAAATACATCCCACTCCCTCCAGAGAGAGTGGCTG 1365  
Db 286 -----ProAsnPheProAlaSerGln----- 292  
QY 1366 GGAACCTCTATTGAGGAATGAGAAAATGCTCGGTGGGATCCATTGCTTCTCTTA 1425  
Db 293 GlyLysLeuArgGluAspAsnArgTyrGlyProLysIleValLeu----- 307  
QY 1426 AGACCATCTGCAAGTCACACAGTACTGTTCGGTAGACCTTCTGAGAGCAGGAGGTT 1485  
Db 307 ----- 307  
QY 1486 CCGAAACCTTTTCCAAACATTAACAAGATTTGTGGGATAACAACAATGTGAAAATGCCT 1545  
Db 308 -----ProGlnArgTrpArgGluPhe----- 314  
QY 1546 TGTTCCGAACAAACTTGTACCTGTGGNAGATGAGATGGTGGAGCA----- 1593  
Db 315 -----AspSerArgGlyArgArgAspSerTyr 324  
QY 1594 -----ACTGCAGG----- 1602  
Db 325 PheTyrPheLysArgLysLeuAspGlyTyrLeuLysCysTyrLysThrThrGlyTyrPhe 344  
QY 1603 -----AGTAGTGGGAGCTCATTCAGACTGCACCTTCTCAACAA 1641

QY 2749 ACTTCTCTACCGAGAGAGCTGGATGTTGGAAAGTGTCACAAAGTTATTGCTTAGATAC 2808  
Db 544 LeuLeuPheArgAsnAspGlyThrValGlnGlnLeuTrpSer---IleLeuArgSer 562  
QY 2809 TAAATGAAGATCCAGAAAGCTGTTCCACCCCTGGACCA-----GAC 2850  
Db 563 TyrSerArgLeuLeuLysGluLysSerSerLysGluProArgGluAsnLysAlaSerLys 582  
QY 2851 ATCAAGCTTTATCCATTCATACCATGCTCTGTCAGTCAAGTCAGAGACCCACTGCATG 2910  
Db 583 LysLysLeuTyArgPheLeu-----LysGluGluLeuLysLysValArgAspVal 599  
QY 2911 CCAGGACAGAGGCA 2925  
Db 600 ProGlyGluGlyAla 604  
RESULT 12  
US-09-511-507-8  
; Sequence 8, Application US/09511507  
; Patent No. 6395543  
; GENERAL INFORMATION:  
; APPLICANT: JACOBSON, Myron K.  
; APPLICANT: JACOBSON, Elaine L.  
; APPLICANT: AME, Jean-Christophe  
; APPLICANT: LIN, Winston  
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE (PARG) EN  
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV  
; TITLE OF INVENTION: THEREWITH  
; FILE REFERENCE: NIA 201  
; CURRENT APPLICATION NUMBER: US/09/511,507  
; CURRENT FILING DATE: 2000-02-23  
; PRIOR APPLICATION NUMBER: 09/302,812  
; PRIOR FILING DATE: 1999-04-30  
; NUMBER OF SEQ ID NOS: 38  
; SEQ ID NO 8  
; LENGTH: 768  
; TYPE: PRT  
; ORGANISM: Drosophila melanogaster  
; FEATURE:  
US-09-511-507-8  
Alignment Scores:  
Pred. No.: 1.49e-94 Length: 768  
Score: 1074.50 Matches: 244  
Percent Similarity: 58.12% Conservative: 96  
Best Local Similarity: 41.71% Mismatches: 194  
Query Match: 15.86% Indels: 51  
DB: 4 Gaps: 15  
US-09-302-812-5 (1-3814) x US-09-511-507-8 (1-768)  
QY 1249 AGAATTTCCAAAGTCAGAGACAGAGAAAGAAACAATGTGAAGTCAGACATCAAGAACA 1308  
Db 45 ArgMetSerLysSerProAspGlyGlyLeuSerGluLeuGluThrGluGluGlu----- 62  
QY 1309 GAAAGGAAGATCCAAATAATCCCACTACCTCCCTCCCA-----GAGAAGAGTGG 1362  
Db 63 -----ProGluAsnLeuAlaAsnSerLeuAspSerTrp 74  
QY 1363 CTGGGAATCTTATGAGAAATGAGAAATGCGCTGCTGGGATCCATTGCTTCC 1422  
Db 75 ArgGlyValSerMetGluAlaIleHisArg---AsnArgGlnProPheGluLeuGluAsn 93  
QY 1423 TTAAGACCA-----TCTGCAAGTCACAGCTGATGTTCCGGGTAGACCTTCTGAGAGCAGGA 1479  
Db 94 LeuProProValThrAlaGlyAsnLeuHisArgValMetTyrGlnLeuProLeuArgGlu 113  
QY 1480 GAGGTTCCGAACCTTTTCCACACATTCACAAAGATTGTGGGATAACAACATGTGAA 1539  
Db 114 ThrProProArgProTyrLysSerProGlyLys-----TrpAspSerGluHisValArg 131  
QY 1540 ATGCTTGTTCGGAACAAACTTGTACCTGTGGGAAGATGAGATGGTGAGCAACTGCA 1599

Db 132 LeuProCysAlaProGluSerLysTyrProArgGluAsnProAspGlySerThrThrIle 151  
QY 1600 GGGAGTAGGTGGAGCTCATTCAGACTGCACTTCTCAACAAATTCACAGACCCAGAAC 1659  
Db 152 AspPheArgTrpGluMetIleGluArgAlaLeuLeuGlnProIleLysThrCysGluGlu 171  
QY 1660 TTGAGATCCGATTCTGAATAATCAATGTCGATATCTAAGAAATGGGACTTTACAGCT 1719  
Db 172 LeuGlnAlaIleIleSerTyrAsnThrThrTyrArgAspGlnTrpHisPheArgAla 191  
QY 1720 TTGCTTCTGCGATAAGTACTTGAAGAAGCAGAGCCCAACATTTATATTCAGTCC 1779  
Db 192 LeuHisGlnLeuLeuAspGluGluLeuAspGluSerGluThrArgValPhePheGluAsp 211  
QY 1780 ATTTACTCAGATGTTGAAATTTGCACTCTGTCGCAATATTTGCAACCCAGCCAAATA 1839  
Db 212 LeuLeuProArgIleIleArgLeuAlaLeuArgLeuProAspLeuLeuGlnSerProVal 231  
QY 1840 CCATCTCCTGAAACAGAGATGATCATCTCTGTCAGATGTCACAGGACAGATCGCCAGT 1899  
Db 232 ProLeuLeuLysHisLysAsnAlaSerLeuSerLeuSerGlnGlnGlnIleSerCys 251  
QY 1900 CTTTCTAGCTAATGCTTCTCTGTCACATTTCCCGACGGAATGCC---AAGATGAAATCG 1956  
Db 252 LeuLeuAlaAsnAlaPheLeuCysThrPheProArgArgAsnThrLeuLysArgLysSer 271  
QY 1957 GAGTATCTAGTTACCCAGACATTAATCAATCGGTTGTTGAGGACGCTTCATCAAG 2016  
Db 272 GluTyrSerThrPheProAspIleAsnPheAsnArgLeuTyrGlnSerThrGlyProAla 291  
QY 2017 AAACGAGAAACTGAAACACATCTCTGCTACTTCTGAAAGAGTC-----ACAGAGAAA 2070  
Db 292 ValLeuGluLysLeuLysCysIleMetHisTyrPheArgArgValCysProThrGluArg 311  
QY 2071 AAA-----CCTACAGGATTTGTCACATTTACAAGACAGAGT-----CTTGAA 2112  
Db 312 AspAlaSerAsnValProThrGlyValValThrPheValArgArgSerGlyLeuProGlu 331  
QY 2113 GATTTCCAGAAATGGGAAAGGTGTGAAAGCTCTG-----ACACGCTTACAGTCACT 2166  
Db 332 HisLeuIleAspTrpSerGlnSerAlaAlaProLeuGlyAspValProLeuHisValAsp 351  
QY 2167 TACGAGGTACCATAGAGGCAACGCGGAGGATGTCACAGTGGATTTTGCAATCGT 2226  
Db 352 AlaGluGlyThrIleGluAspGluGlyIleGlyLeuLeuGlnValAspPheAlaAsnLys 371  
QY 2227 TTTCTTGGAGTGTGTGACTGTCGCGGACTTGTACAAGAAGAAATCAGATTTTAATC 2286  
Db 372 TyrLeuGlyGlyValLeuLeuGlyHisGlyCysValGlnGluGluIleArgPheValIle 391  
QY 2287 AATCCTGAATGATGTTTACGGCTGTTTACGGCTGTTTACGGCTGTTGATCACAATGATGCTT 2346  
Db 392 CysProGluLeuLeuValGlyLysLeuPheThrGluCysLeuArgProPheGluAlaLeu 411  
QY 2347 ATTATCACAGTACTGAAACAGTACAGAGTAATACACAGGCTATGCTGAAACTTATCGTTGG 2406  
Db 412 ValMetLeuGlyAlaGluArgGlySerAsnTyrThrGlyThrGlyThrGlyThrGlyThrGly 431  
QY 2407 GCCCGAAGCCATGAAGATGGAGTGAAGAGCAGCATTTGGCAGCGGCTGCTGACAGAGATC 2466  
Db 432 SerGlyAsnPheGluAspSerThrProArgAspSerSerGlyArgArgGlnThrAlaIle 451  
QY 2467 GTTGCCATTCAGCAGCTTCTCAGCGCTACCTCGATCAGTTGTCGCTGAGAAAGTG 2526  
Db 452 ValAlaIleAspAlaLeuHisPheAlaGlnSerHisHisGlnTyrArgGluAspLeuMet 471  
QY 2527 AGAGCTGAGCTTACAAAGGCTTACTCGGATTCCTCCGT-----CCT 2568  
Db 472 GluArgGluLeuAsnLysAlaTyrIleGlyPheValHisTrpMetValThrProProPro 491  
QY 2569 GGAGTCTCTCTGAAATCTTCTGCTGAGTGGCCAGGAAACTGGGCTGTGTGCTCTT 2628

QY 2851 ATCAAGCTTTATCCATTATATACCATCTGTTGAGTCAAGTGCAGAGACCACTCAGATG 2910  
 |||||  
 Db 583 LysLysLeuTyAspPhele-----LysGluGluLeuLysLysValArgAspVal 599  
 |||||  
 QY 2911 CCAGCAGCAGAGGCA 2925  
 |||||  
 Db 600 ProGlyGluGlyAla 604  
 |||||  
 RESULT 11  
 US-09-511-477-8  
 ; Sequence 8, Application US/09511477  
 ; Patent No. 6337202  
 ; GENERAL INFORMATION:  
 ; APPLICANT: JACOBSON, Myron K.  
 ; APPLICANT: JACOBSON, Elaine L.  
 ; APPLICANT: AME, Jean-Christophe  
 ; APPLICANT: LIN, Winston  
 ; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE (PARG) EN  
 ; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV  
 ; TITLE OF INVENTION: THERWITH  
 ; FILE REFERENCE: NIAD 201  
 ; CURRENT APPLICATION NUMBER: US/09/511,477  
 ; CURRENT FILING DATE: 2000-02-23  
 ; PRIOR APPLICATION NUMBER: 09/302,812  
 ; PRIOR FILING DATE: 1999-04-30  
 ; NUMBER OF SEQ ID NOS: 38  
 ; SEQ ID NO 8  
 ; LENGTH: 768  
 ; TYPE: PRT  
 ; ORGANISM: Drosophila melanogaster  
 ; FEATURE:  
 US-09-511-477-8

Alignment Scores:  
 Pred. No.: 1,49e-94 Length: 768  
 Score: 1074.50 Matches: 244  
 Percent Similarity: 58.12% Conservative: 96  
 Best Local Similarity: 41.71% Mismatches: 194  
 Query Match: 15.86% Indels: 51  
 DB: 4 Gaps: 15

US-09-302-812-5 (1-3814) x US-09-511-477-8 (1-768)

QY 1249 AGAATTCAGTCAAGCAGAGAGAGAAAGCAATGTGAAGTCAGACATCAAGAGACA 1308  
 |||||  
 Db 45 ArgMetSerLysSerProAspGlyGlyLysSerGluLeuGluThrGluGluGlu----- 62  
 |||||  
 QY 1309 GNAAGAGATTCCAAATACATCCACCTACCTCCCTCCA-----GAGAGAGAGTGG 1362  
 |||||  
 Db 63 -----ProGluAsnLeuAlaAsnSerLeuAspSerTirp 74  
 |||||  
 QY 1363 CTGGGAACCTCTATTGAGGAATGAGAAATGCCTCGGTGGGATCCATTGGCTTCC 1422  
 |||||  
 Db 75 ArgGlyValSerMetGluAlaIleHisArg---AsnArgGlnProPheGluLeuGluAsn 93  
 |||||  
 QY 1423 TTAAGACCA---CTGCAAGTCACAGTCACTGTTCGGGTAGACCTTCTGAGAGAGGA 1479  
 |||||  
 Db 94 LeuProProValThrAlaGlyAsnLeuHisArgValMetTyrGlnLeuProIleArgGlu 113  
 |||||  
 QY 1480 GAGGTTCCGMAACCTTTCCACACATTAACAAGATTTGTGGATGAACAACATGTGAA 1539  
 |||||  
 Db 114 ThrProProArgProTyrLysSerProGlyLys-----TirpSerGluHisValArg 131  
 |||||  
 QY 1540 ATGCTTGTTCGGAACAAACTGTACCTGTGGAGATGAGATGGTGAGCGAACTGCA 1599  
 |||||  
 Db 132 LeuProCysAlaProGluSerLysTyrProArgGluAsnProAspGlySerThrThrIle 151  
 |||||  
 QY 1600 GGGAGTAGTGGAGCTCATTCAGTGCACCTCTCAACAATTCACAGACCCCGAAC 1659  
 |||||  
 Db 152 AspPheArgTirpGluMetIleGluArgAlaLeuLeuGlnProIleLysThrCysGluGlu 171  
 |||||  
 QY 1660 TTGAGGATGCCATTCGAAATACAAATGTGGCATATTCAGAAATGGGACCTTACAGCT 1719  
 |||||

Db 172 LeuGlnAlaAlaIleLeuSerTyrAsnThrThrTyrArgAspGlnTirpHisPheArgAla 191  
 QY 1720 TTGGTTGATTCTTGGGATATAGGTACTTGAAGACAGAGAGGCCAACATTTATATCAGTCC 1779  
 |||||  
 Db 192 LeuHisGlnLeuLeuAspGluGluLeuAspGluSerGluThrArgValPhePheGluAsp 211  
 |||||  
 QY 1780 ATTTTACCTGACATGGTGAATAATTGCCTCTGTCTGCCAAATATTTGACCCAGCCAATA 1839  
 |||||  
 Db 212 LeuLeuProArgIleAlaLeuAlaLeuArgLeuProAspLeuLeuIleGlnSerProVal 231  
 |||||  
 QY 1840 CCACCTCTGAAACAGACAGATGAATCATCTGTGACGATGTACAGAGACAGATCCCACT 1899  
 |||||  
 Db 232 ProLeuLeuLysHisHisLysAsnAlaSerLeuSerGlnGlnGlnLysSerCys 251  
 |||||  
 QY 1900 CTTTGTAGCTAATGCTTTCTTCGCACTTCCCGACGGAATGCC---AAGATGAATAATCG 1956  
 |||||  
 Db 252 LeuLeuAlaAsnAlaPheLeuCysThrPheProArgAsnThrLeuLysArgLysSer 271  
 |||||  
 QY 1957 GAGTATTCTAGTACCCAGACATTAACCTCAATCGTGTGTGAAGGACGTTTCATCAAGG 2016  
 |||||  
 Db 272 GluTyrSerThrPheProAspIleAsnPheAsnArgLeuTyrGlnSerThrGlyProAla 291  
 |||||  
 QY 2017 AAACCCAGAAAAAAGTGAACACTCTCTGCTACTTTTCAAGAGATC-----ACAGAGAAA 2070  
 |||||  
 Db 292 ValLeuGluLysLeuLysCysIleMetHisTyrPheArgArgValCysProThrGluArg 311  
 |||||  
 QY 2071 AAA-----CCTACAGGATTTGTGACATTTTACAAGACAGAT-----CTTGA 2112  
 |||||  
 Db 312 AspAlaSerAsnValProThrGlyValValThrPheValArgArgSerGlyLeuProGlu 331  
 |||||  
 QY 2113 GATTTCCAGATGGAAAGGTGTGAAGGCTCTG-----ACAGCTTACACCTCACT 2166  
 |||||  
 Db 332 HisLeuIleAspTirpSerGlnSerAlaAlaProLeuGlyAspValProLeuHisValAsp 351  
 |||||  
 QY 2167 TACGAGGCTACCATAGAAGGCAACCGCGAGCATGCTACAGTGGATTTGCAATCGT 2226  
 |||||  
 Db 352 AlaGluGlyThrIleGluAspGluGlyIleGlyLeuLeuGlnValAspPheAlaAsnLys 371  
 |||||  
 QY 2227 TTTGTTGAGGTGTGTGACTGTCTCGGGACTTGTACAAGAGAAATCAGATTTTAAATC 2286  
 |||||  
 Db 372 TyrLeuGlyGlyGlyValLeuGlyHisGlyCysValGlnGluIleArgPheValIle 391  
 |||||  
 QY 2287 AATCTGATGATGTTTTCACGGCTGTCTAGGTGCTGCATCACAATCAGTGTCTT 2346  
 |||||  
 Db 392 CysProGluLeuLeuValGlyLysLeuPheThrGluCysLeuArgProPheGluAlaLeu 411  
 |||||  
 QY 2347 ATTATCACAGGTACTGAACAGTACAGTGAATACACAGGCTATGTGAAACTTATCGTTGG 2406  
 |||||  
 Db 412 ValMetLeuGlyAlaGluArgTyrSerAsnTyrThrGlyTyrAlaGlySerPheGluTirp 431  
 |||||  
 QY 2407 GCCCGAAGCCATGAGATGGGAGTGAAGGACGATTTGGCAGCGGCTGCACGGAGATC 2466  
 |||||  
 Db 432 SerGlyAsnPheGluAspSerThrProArgAspSerSerGlyArgArgGlnThrAlaIle 451  
 |||||  
 QY 2467 GTTGCCATTGACGACCTTCACTTCAGACGCTACCTCGATCAGTGTTCGCTGAGAAAGTG 2526  
 |||||  
 Db 452 ValAlaIleAspAlaLeuHisPheAlaGlnSerHisHisGlnTyrArgGluAspLeuMet 471  
 |||||  
 QY 2527 AGACGTGAGCTTAAACAGGCTTACTGCGGATTCCTCCCT-----CCT 2568  
 |||||  
 Db 472 GluArgGluLeuAsnLysAlaTyrIleGlyPheValHisTirpMetValThrProProPro 491  
 |||||  
 QY 2569 GGAGTTCCTTCTGAAATCTTTCTGTCAGTGGCCAGCGGAAACTGGGGCTGTGGTCCCTTT 2628  
 |||||  
 Db 492 Gly-----ValAlaThrGlyAsnTirpGlyCysGlyAlaPhe 503  
 |||||  
 QY 2629 GGGGTGAGCTAGATTAAGACCTTAATACAGATCTCGGACGCTGCTGGCGGCTGAAGCT 2688  
 |||||  
 Db 504 GlyGlyAspSerTyrLeuLysAlaLeuLeuGlnLeuMetValCysAlaGlnLeuGlyArg 523  
 |||||  
 QY 2689 GACGTGGTTCATTCACCTTTGGGACCTCAGAGTTGTAGAGACATTTACAGCATGCAC 2748  
 |||||  
 Db 524 ProLeuAlaTyrTyrThrPheGlyAsnValGluPheArgAspPheHisGluMetTirp 543  
 |||||



QY	787	GAGAGTCCTTTGTCAGATGTTGGTCCGAGGACATTTGGAACTGGACCAAAAAATGACAAC	846
Db	261	GlusErProLeuSerAspThrGlySerGluAspValGlyThrGlyLeuLysAsnAlaAsn	280
QY	847	AAATTCAGTCGACAGAAAGCAGCCCTAGCTGATTCGCTCCCTCATTTTGAGAAAGAAAGTGAG	906
Db	281	ArgLeuAsnAaGgInGluSerSerLeuGlyAasnSerProProPheGluLysGluSerGlu	300
QY	907	CCTGAGTCACCAATGGATGTGTAGACAACCTCGAGAAACAGATTGTCAGAGTTCAGAAAGCAGAT	966
Db	301	ProGluSerProMetAspValAspAsnSerLysAasnSerCysGlnAaspSerGluAlaAasp	320
QY	967	GAAGAAACAATCCAGTCTTTTCATGACGACGACAGATGATCGTTCCTCC---CAAACAGCAAAAT	1023
Db	321	GluGluThrSerProGlyPheAaspGluGlnAaspSerSerSerAlaGlnThrAlaAsn	340
QY	1024	AAACTTTTCAAGTTGCCAAGCAAGAGAAGCTGATCGCATCTTAGCAACCGTATTTCGACT	1083
Db	341	LysProSerAaPheGlnProArgGluAlaAspThrGluLeuAaGlyLysAaGSerSerAla	360
QY	1084	AAGGGAAGTGAAGTTAGATTGCATTTCCAAATTCGAA---GGAGAAAAATAATGCTGGGACC	1140
Db	361	LysGlyGlyGluLeuAaGLeuHisPheGlnPheGluGlyGlyGluSerAaGAlaGlyMet	380
QY	1141	AGTGACTTAATGTCGACCAACCCATCTGGAAACTCTTCTAGCCTTATGTAGAGTGTAGAGT	1200
Db	381	AsnAaspValAsnAlaLysAaGProGlySerThrSerSerLeuAsnValGluCysAaGAsn	400
QY	1201	TCCAAGCAGCATGGAAAAAGGAGTTCTAAAAATTACAGATCATTTTCATGAGAAATTTCCAAG	1260
Db	401	SerLysGlnHisGlyAaGlyAaSerLysIleThrAspHisPheMetAaGValProLys	420
QY	1261	TCAGAGCAGACAGAAAGAAACAATGTGAGTCCAGATCAATCAAGAAACAGAAAGAGAGATT	1320
Db	421	AlaGluAspLysAaGlyLysGluGlnCysGluMetLysHisGlnAaGThrGluAaGlySle	440
QY	1321	CCAAAAATACATCCACCTTAACCTCCCTCCAGAGAGAGAGTGGCTGGGAACCTCTATTTGAG	1380
Db	441	ProLysTyrIleProProHisLeuSerProAaspLysLysTrpLeuGlyThrProIleGlu	460
QY	1381	GAATGAGAAAAATGCTCGGTGGGATCCATTTGGCTTCCTTAAGACCATCTGCAAGT	1440
Db	461	GluMetAaGArgMetProAaGProCysGlyLysAaGLeuProLeuAaGProSerAlaAsn	480
QY	1441	CACACGTGACTGTTTCGGGTAGACCTTCTCGAGACGAGAGAGTTCGAAACCTTTTCCA	1500
Db	481	HisThrValThrIleAaGValAspLeuLeuAaGlyLeuValProLysProPhePro	500
QY	1501	ACACATTCACAAAGATTTGCGGATAACAAACATGTGAATGCTTGTTCGGNACAAAC	1560
Db	501	ThrHisPheLysAaspLeuTrpAaspAsnLysHisValLysMetProCysSerGluGlnAsn	520
QY	1561	TTGTACCTCTGGAAGATGAGATGGTGACGCAACTCGAGGAGTAGTGGGAGCTCAT	1620
Db	521	LeuTyrProValGluAaspGluAaGnglyLysAaGlySerAaGlySerAaGlyTrpGluLeuIle	540
QY	1621	CAGACTGCATCTTCAACAAATTCACAGACCCAGACATTGAAGATGCGATCTCGAAA	1680
Db	541	GlnThrAlaLeuLeuAsnAaGLeuThrAaGProGlnAsnLeuLysAaspAlaIleLeuLys	560
QY	1681	TACAATGTGCATATTTCAAGAAATGGACTTTACAGCTTTGGTTGATTTCTGGGATAG	1740
Db	561	TyrAsnValAlaTyrSerLysLysTrpAaspPheThrAlaLeuIleAaspPheTrpAaspLys	580
QY	1741	GTACTTTGAAGAGCAGAGGCCCAACATTTATATATATATATATATATATATATATATATAT	1800
Db	581	ValLeuGluGluAlaGluAlaGlnHisLeuTyrGlnSerIleLeuProAaspMetValLys	600
QY	1801	ATTGCACTGTCTGCCAATATTTGACCCAGCCCAATACCACTCTCTGAAAACAGAGATG	1860
Db	601	IleAlaLeuCysLeuProAsnIleCysThrGlnProIleProLeuLeuLysGlnLysMet	620

RESULT 10